Endogenous Firm Structure and Worker Specialization

Jacob Adenbaum*

February 2023

CLICK HERE FOR MOST RECENT VERSION

Abstract

What tasks must be performed to produce a good? Which occupations are well suited to do those tasks? And what are the gains to worker specialization within the firm? I use Brazilian administrative data to document new facts about how firms vary the types of workers that they choose to hire as they grow larger. Bigger firms hire more distinct occupations. They also hire a set of workers whose cognitive, manual, and interpersonal skills are more dispersed than at small firms. I then develop a structural model of how firms choose which types of workers to hire, and how they assign tasks to these workers. I propose a novel identification strategy for how to indirectly infer the (multi-dimensional) distribution of skill requirements for tasks that firms face and using only cross-sectional data on which occupations firms choose to hire, and in what proportion, across the firm size distribution. I estimate my model using Brazilian manufacturing firms, and show that more than 1/3 of the variance in firm level TFP is due to firms' endogenous choices of which types of workers to hire (and how specialized those workers should be). I find that gains from increasing firm specialization are about 1.3% of output, and that the costs to shutting down worker specialization within firms are large, leading to a 9.6% decrease in total output. I find similar gains in more narrowly defined industry codes such as leather goods.

Keywords: Occupational Choice, Skills, Firm Structure

JEL Codes: J23, J24, L23

*University of Edinburgh

1 Introduction

Economists have long understood that the productivity of workers depends on how they are organized in production. A worker's output depends not only on their skills, but on the specific tasks that they are assigned to do in production, and how well suited they are to do those tasks. However, which tasks they are assigned depends, crucially, on the composition of their coworkers. In a small grocery store, a single worker may stock the shelves, sweep the floors, staff the cash register, *and* plan the inventory orders. However, in a larger grocery chain, these roles may be split up into separate occupations: cashiers who are dedicated to checking out customers and stocking the shelves, janitors who focus on keeping the shop clean, and managers who plan the store's inventory from week to week. The set of tasks that workers perform, and therefore their productivity, depends on the firm's scale of production and the degree of specialization that it can support.

In this paper, I seek to quantify, at the aggregate level, the gains from worker specialization within firms. I focus on three questions: What tasks must be performed to produce a good? Which occupations are well suited to do those tasks? And what are the productivity gains from reorganizing firms to use the optimal mix of occupations to complete these tasks?

To answer these questions, I develop a new theoretical model of how firms choose both the set of occupations to employ in production, and how to assign tasks, with heterogenous skill requirements, to workers in those occupations. My paper contributes a new theory of firm's endogenous choice of their organizational structure, reminiscent of the work of Garicano and Rossi-Hansberg (2006), but allowing for multiple dimensions of skill (as in Ocampo (2019), Lindenlaub (2017)) and heterogenous firms. I estimate my model using administrative data on manufacturing firms in Brazil, and use the results to quantify the gains from specialization. This measurement strategy exploits the detailed information on the occupations of the workers hired by firms available in the Brazilian administrative data. Such datasets have only become available to researchers in recent years.

In the model, heterogeneous firms must choose how to optimally allocate a discrete distribution of tasks to a finite number of worker types, which correspond to the various worker types that firms choose to hire. Worker types are a bundle of (multi-dimensional) skills, and tasks are defined by their skill requirements.[1]

I extend the task assignment framework to allow for productivity differences across firms. Firms choose how many types of workers to hire, as well as the skill bundle that each worker type will have. Firms pay an organizational cost for each additional type of worker that they choose to hire, which captures the additional organizational and managerial overhead involved in managing more types of workers. This cost leads firms to endogenously choose to hire only a finite number of worker types. Firms take the wage function (i.e, the mapping between worker skills and the wage they must pay) as given. In this environment, firms face two main choices when organizing production: Firms choose how to assign tasks to workers, as well as which set of occupations to hire (and in what quantities).

Task Assignment Problem. The firm must complete a pre-specified set of tasks to produce. These tasks come in fixed proportions, which the firm takes as given. In the task assignment problem, firms choose how to assign each task to a worker, subject to feasibility constraints. That is, they cannot assign a worker more tasks than they have the time to complete, nor can they assign a worker tasks that do not exist. A paralegal in a law firm has only so many hours in the day, but also, there are only so many letters that need to be opened. As in Ocampo (2019), this task assignment problem takes the form of an optimal transport problem, which allows for convenient analytic results, and lends itself well to efficient numerical solution methods as it is a special type of linear program.

Occupational Choice Problem. Firms face a tradeoff between their degree of worker specialization and the organizational cost of hiring and managing more types of workers. At

[1]The model builds on Ocampo (2019), who considered the multi-dimensional task assignment problem from the perspective of a single production unit. I allow for heterogeneous firms, which is key to identifying the primitives of the model using linked employer-employee data.

one extreme, firms might hire many different types of workers, who will all perform tasks that are narrowly tailored to their skills. By minimizing worker-task mismatch, these firms can achieve productivity gains. However, they must pay a higher organizational cost to manage all of these different roles within the firm. At the other extreme, a firm might choose to hire just a single worker type, who will perform all of the necessary tasks. These workers will be less productive because of the high degree of mismatch between their skills and the tasks they are asked to do, but the firm will economize on organizational costs. Firms have an optimal scale that balances these two competing concerns, and higher productivity firms (for whom the organizational overhead is relatively less important) will endogenously choose a higher degree of worker specialization. This relationship between firm productivity and organizational complexity is similar to the unidimensional results theorized and documented in Rosen (1982), Garicano (2000) and Caliendo, Monte, and Rossi-Hansberg (2015). I allow for tasks and workers to be heterogenous along several dimensions of skill, which is important to match the empirical patterns of specialization that we observe in the data.

One of the main contributions of this paper is to show how to estimate a task assignment model of this kind using matched employer-employee data. Many other papers on worker specialization within the firm, such as Ales, Combemale, Fuchs, and Whitefoot (2021), have relied upon detailed industry surveys and hand-collected data on the tasks that firms use in production. As a result, it is difficult to use these approaches to speak to the macroeconomic implications of worker specialization. My novel identification strategy, which exploits cross-sectional heterogeneity across different sized firms allows me to explore the role of worker specialization in driving *aggregate* productivity. I show how to recover estimates of the distribution of tasks and of the worker-task production function using only a single cross-section of firms.

The key difficulty here is that it is not straightforward to infer the types of tasks that workers perform within a firm simply by observing the skills of the workers because within a firm, workers are substitutes for one another. Although it may not be the best use of their time, a highly paid attorney could sort mail and answer phones. In fact, at small law offices, they often do. A cashier at a small store might also sweep the floors, whereas a larger retail business would hire a separate janitor.

To infer what kinds of tasks firms perform, one must take careful account of the fact that workers with similar skills will perform *different* tasks within the firm depending on the skills of their coworkers. It is this fundamental identification problem that necessitates a structural model of how firms assign tasks to workers, and how they choose which workers to hire, in order to make sense of the data we observe.

To make progress on this, I develop a novel identification strategy that allows me to estimate the distribution of tasks that firms face by using data on which occupations firms choose to hire, and in what shares, across the distribution of firm sizes. When the firm behaves optimally, they choose a labor allocation which equates each occupation's share of the total wage bill to that occupations share of total output. I target the wage bill shares in my estimation strategy, and show that we can use these moments to recover the primitives of the task-based production function. Crucially, I observe workers with similar skills in firms with *different* configurations of coworkers. So long as they are assigned to a different set of tasks, this provides new information about what tasks the firms perform.

I implement this estimation strategy using administrative linked employer-employee data from Brazil: the Relação Anual de Informações Sociais (RAIS). I interpret the worker types from the model as worker occupations. To measure worker skills, I calculate measures of occupational skills in the cognitive, manual, and social dimensions from O*NET, which I construct as in Lise and Postel-Vinay (2020) and then merge into the RAIS.

I then use my estimates (for the manufacturing sector) to quantify the aggregate contribution of worker specialization within the firm to observed productivity, and to evaluate the welfare gains from increasing the degree of specialization within firms. I find that shutting down the within firm endogenous specialization channel leads to aggregate output losses of 9.6%. Moreover, I show that this endogenous worker specialization accounts for 36% of the variation in firm-level TFP. I also find limited, but not insubstantial potential gains to increasing the degree of worker specialization within the firm, which are around 1.3% of total sector output.

I show that these results are robust to more narrow specifications of industry by re-estimating the model for more specialized industries that specialize in the production of homogeneous goods, such as the leather working industry. I show that these industries exhibit similar patterns of firm specialization as for the aggregate manufacturing sector.

This paper proceeds in several sections. In section 2, I review the related literature. Section 4 describes the model I use, and presents several important analytic results. Section 3 describes the RAIS and O*NET data that I use to estimate the model, and provides descriptive results, and Section 5 describes the details of my estimation procedure, and presents the results. I conclude in Section 7.

2 Related Literature

This paper builds on the existing literature that uses task assignment models to model the firm's choice to hire heterogeneous labor inputs in production. Many papers have adopted a task based approach to modeling the firm's production technology (for instance, Rosen (1978), Acemoglu and Autor (2011)). These task assignment models treat worker skills as being unidimensional. Like Lindenlaub (2017), I allow for several dimensions of skill. This allows me to capture important heterogeneity in how the skills hired by firms vary across the distribution of firm size.

My work is most closely related to Ocampo (2019), which also tackles the problem of assigning workers to tasks in several skill dimensions by formulating the assignment problem as an optimal transport problem (Villani, 2009; Galichon, 2016), as in Lindenlaub (2017). I build on this theory in two main ways: First, I extend the theory to accommodate heterogenous firms, whose choice of *which* occupations they want to hire yields an endogenous hierarchy of specialization across firms. More productive firms choose to hire a more specialized set of workers in production and divide the tasks among them more efficiently to realize lower costs of production. Second, observing workers with identical skills being employed in different production configurations (with a different set of coworkers) provides me with a rich source of cross-sectional variation that allows me to make substantial progress in identifying the distribution of tasks that firms must complete in order to produce. I use the moments implied by the firms' first order conditions to estimate the task distribution, and provide a constructive proof of identification.

An alternate approach to measuring the actual tasks that firms must complete has been proposed by Ales, Combemale, Fuchs, and Whitefoot (2021). They exploit detailed surveys of plant-level data, paired with a detailed task assignment model to recover estimates of the firm's production function. My approach complements this strategy; by allowing for multiple dimensions of worker skill, and relying on O*NET measures of those skills, I estimate my model using administrative data based on all firms in the economy, which allows me to measure the aggregate impacts of the endogenous specialization channel.

This paper also contributes to our understanding of the firm's endogenous choice of its organizational structure. Previous theoretical work on this problem, in Rosen (1982), Garicano (2000), and Garicano and Rossi-Hansberg (2006) generates an endogenous organizational hierarchy *within* the firm by assuming a production technology where workers, who differ only along a single dimension of skill, pass tasks which are too difficult for them to complete up to their managers. I contribute to this literature by providing an alternate approach to modeling the firm's organizational problem, which allows for multiple dimension of worker skill. In my context, I do not impose the hierarchical structure of production a priori. Rather, firms are faced with the choice of which occupations they want to hire, subject to occupation specific fixed costs; these fixed costs play a similar role to the information processing frictions which proved crucial in earlier work, while remaining agnostic about the specific hierarchy of the firm's reporting structure.

The main empirical findings of my paper, how the number of workers and their skills vary with the firm size distribution, contribute to the empirical literature on how the firm's internal organization varies with size. Other papers, such as Caliendo, Monte, and Rossi-Hansberg (2015), Caliendo, Mion, Opromolla, and Rossi-Hansberg (2020), Tåg (2013), and Friedrich (2020) have studied the organizational structure of firms using administrative data. Although these papers focus on management layers rather than occupations, my empirical findings are broadly in line with their results. This paper's main contribution, relative to that literature, is to document that the average skills hired by firms vary non-monotonically with firm size, and that this varies substantially by skill type (among cognitive, manual, and interpersonal skills).

3 Data

I make use of two datasets for my analysis. First, I use the Brazilian *Relação Anual de Informações Sociais (RAIS)*, a rich and high quality administrative panel dataset covering the entire Brazilian formal sector which measures worker earnings and demographic characteristics, matched to the firms that employ them. Second, I merge into the RAIS occupational measures of skill from O*NET, in order to associate with each worker a measure of their cognitive, manual, and interpersonal skills.

In this section, I will give an overview of both of these datasets, and document several novel facts about how the types of skills employed by Brazilian firms varies with firm size.

3.1 Matched Employer-Employee Data (RAIS)

The RAIS is an administrative matched employer-employee dataset collected by the Brazilian *Ministerio da Economia* (Ministry of the Economy).[2] Every Brazilian firm that has a tax identification number (*Cadastro Nacional de Pessoa Jurídica* — CNPJ) must file an annual report to the government detailing the workers they hired and the wages they paid. Regulatory compliance is enforced through a deterrence mechanism of fines and periodic audits.

The RAIS has near universal coverage of all workers employed in Brazil's formal sector, starting in 1994. It is worth noting that, like many Latin American countries, Brazil has a relatively large *informal* sector, which the RAIS does not cover.[3] Because of these data limitations, my analysis is restricted to formal sector-firms.

My sample covers the time period from 1994 to 2010, and includes a rich set of measures of worker and firm characteristics. I observe anonymized firm, establishment, and worker identifiers. These allow me to track workers and firms across time. Firms report, for every worker they hire, the start and end months of the job spell, the weekly contracted hours, the worker's monthly earnings, as well as a set of demographic variables including age, sex, nationality, education, and race.

Crucially, I observe both industry and occupation codes at the 5-digit level using the Brazilian CNAE (*Classificação Nacional de Atividades Económicas*), and CBO (*Classificação Brasileira de Ocupações*) respectively. The universal reporting of occupation codes is a relatively unusual feature of the Brazilian administrative data, which I exploit to obtain occupation-specific measures of worker skill.

[2] Until 2019, it was collected by the *Ministerio de Trabalho e Emprego* (Ministry of Labor and Employment), which was subsumed into the newly founded Ministry of the Economy as part of a government-wide reorganization and administrative consolidation under the Bolsonaro administration.

[3] Dix-Carneiro, Goldberg, Meghir, and Ulyssea (2021) find that approximately 48% of the labor force is employed in the informal sector, and that this comprises nearly two thirds of firms. The formal sector, however accounts for the majority (60%) of total output.

I drop from my sample workers who are missing earnings or occupation data, as well as the top and bottom 5% of wage earners, but I do not impose any other sampling restrictions. This is because my analysis relies, as much as possible, on observing the full composition of each firm's workforce and their skills.

3.2 Occupational Skill Data (O*NET)

I use the Occupational Information Network (O*NET) Database to generate occupation specific measures of worker skills. O*NET collects detailed survey data on occupations by surveying businesses (in the US) about the characteristics of the occupations that they employ.

O*NET collects detailed survey data on 970 different occupations in the US, with 277 distinct descriptors. Each descriptor includes information about its relative importance, the level, and the relevance to the job. To extract measures of occupational skill from this data, I follow the procedure in Lise and Postel-Vinay (2020). That is, I use a singular value decomposition to extract the first three principal components of the skill measures. To obtain measures that are interpretable as cognitive, manual, and interpersonal skill, I choose three skill measures which I will interpret as reflecting only one of the three dimensions of skill. I then find the linear combinations of the first three principal components that satisfy three exclusion restrictions:

1. The *mathematics* score reflects only cognitive skills. That is, it is orthogonal to my measures of manual and interpersonal skill.
2. The *mechanical knowledge* score reflects only manual skills. It is orthogonal to the measures of cognitive and interpersonal skill.
3. The *social perceptiveness* score reflects only interpersonal skills. That is, it is orthogonal to my measures of cognitive and manual skill.

I then rescale these measures so that they lie in $[0, 1]$. This procedure allows me to generate measures of cognitive, manual, and interpersonal skill that rely on a minimal set of exclusion restrictions, and does not require me to classify the O*NET descriptors as reflecting any one type of skill a priori (other than the three descriptors used in the exclusion restrictions). I map these skill measures to the US Census 2000 occupation codes using the mapping from Sanders (2012).

The O*NET data is unusual for the degree of detail with which it describes the skills and characteristics of workers in each occupation. To the best of my knowledge, comparable surveys have not been conducted specifically in Brazil.[4] To interpret the O*NET survey in the Brazilian context, I rely on the fact that, as Muendler, Poole, Ramey, and Wajnberg (2004) argue, the Brazilian occupational classification system (CBO) is based on a similar set of organizational principles to that of the 1988 International Standard Classification of Occupations (ISCO-88) system. This allows for a reasonable mapping of Brazilian occupations to the US Census codes[5], and for a meaningful interpretation of O*NET occupational skill measures within the Brazilian context.

3.3 Defining effective labor

To measure how firms change the types of workers that they hire, I first need to define a notion of firm size. My preferred measure is a measure of each firm's total quantity of effective labor hired (this is also a measure that is consistent with the model). This analysis is robust to defining firm size in terms of total hours of labor hired, total wage bill, or total number of workers hired.

For each worker i, denote their occupation $n_i \in 1, \ldots, N$, the firm they are employed at as $j_i \in 1, \ldots, J$, and their average monthly earnings as w_i over their employment spell at

[4] The World Bank's STEP Skills Measurement Program is the closest other source of occupational skill data in the international context, but has not yet expanded their data collection efforts to Brazil.

[5] The mapping I use is built off of a crosswalk that was generously provided to me by Gustavo de Souza (See De Souza (2020)).

firm j_i. For every occupation n, let $\overline{w}_n$ denote the average monthly earnings for workers in occupation n, weighted by total hours worked l_i I define worker i's effective labor supplied as:

$$M_i := l_i \times \left(\frac{w_i}{\overline{w}_{n_i}} \right)$$

That is, I adjust the total number of hours worked by the ratio of the worker's wages to their occupation's average wages, in order to get a measure of the worker's efficiency units of labor supplied. I rank the firms by the total quantity of effective labor that they hire, and for each firm j, I compute their percentile rank r_j.

3.4 Stylized Facts

I document several novel facts about how firms vary both the overall number of occupations and the skills of those occupations, all of which suggest that firms are increasing the degree of worker specialization as they get larger. For each percentile rank r, I compute both the average effective labor hired, and the average number of distinct occupations hired (at the 2-digit level), and plot the relationship between them in Figure 1. We see that, for all but the very small firms in the economy (who hire less than a single full-time equivalent worker's worth of effective labor), there is a stark and increasing log-linear relationship between the total quantity of labor hired, and the number of occupations that they choose to employ, with a slope of approximately 0.37. The relatively flat slope among the very small firms may reflect labor indivisibilities (it may be difficult to hire labor in extremely small bundles, so in firms that hire very little labor, it is more difficult to divide labor among several types of workers, since each worker would only be hired for a few hours in a given week).

This relationship, while not necessarily surprising, captures an important dimension of how firms increase the degree of specialization among their workers as they grow. It is not, however, purely mechanical.[6] Crucially, it suggests that firms are increasing the degree of

[6]To see this consider two extreme cases. First, imagine that occupations were entirely uninformative, and were essentially a random labelling. This is the case we would expect to see if *everything* were being

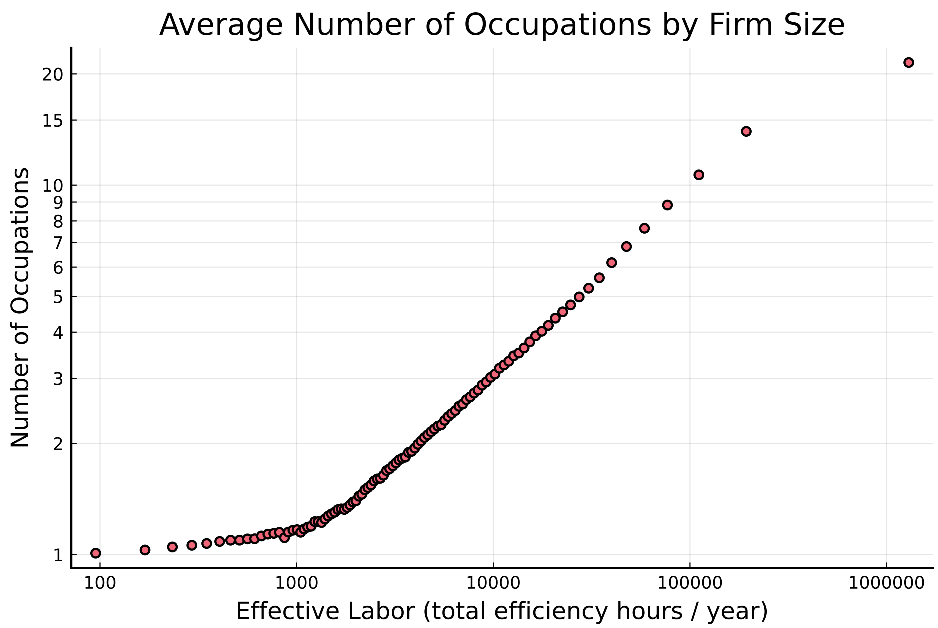


Figure 1: Average number of distinct occupations by firm size, in 2000. Firms are ranked by the total number of efficiency units of labor that they hire in each year, and we compute, for each percentile bin, the average quantity of labor hired in that bin, and the average number of distinct 2-digit occupations hired by firms in that bin. Both axes are on log-scale.

specialization of workers within the firms as they get larger.

However, for understanding worker specialization within the firm, the number of distinct occupations is a relatively coarse measure. Two occupations may be labelled separately, despite having nearly identical skills, and being highly substitutable for one another in production, or may truly reflect distinct skills that are used differently in production. To distinguish between these cases, I will exploit the continuous measures of occupation skill, along the cognitive, manual, and interpersonal dimensions, which are derived from the O*NET

driven by the indivisibility of labor bundles. In this case, we would expect to see far more occupations being hired at comparably sized firms, with a slope of the relationship between the total quantity of labor hired and the number of distinct observations being much closer to 1. In the data, firms increase the number of occupations that they hire much *more slowly* than we would expect to see if occupations were assigned randomly (This is similar to the "Balls in Bins" mechanism explored in Armenter and Koren (2014)).

In the second case, imagine that labor is perfectly divisible among occupations, and firms face no overhead or fixed costs associated with increasing the number of worker types that they choose to hire. In this case, there is an efficient way to divide labor among the various occupations, and every firm chooses exactly the optimal mix of occupations (and therefore the same number of distinct occupations) regardless of size.

These two examples make clear that the slope of this relationship is highly informative about the organizational costs that firms face.

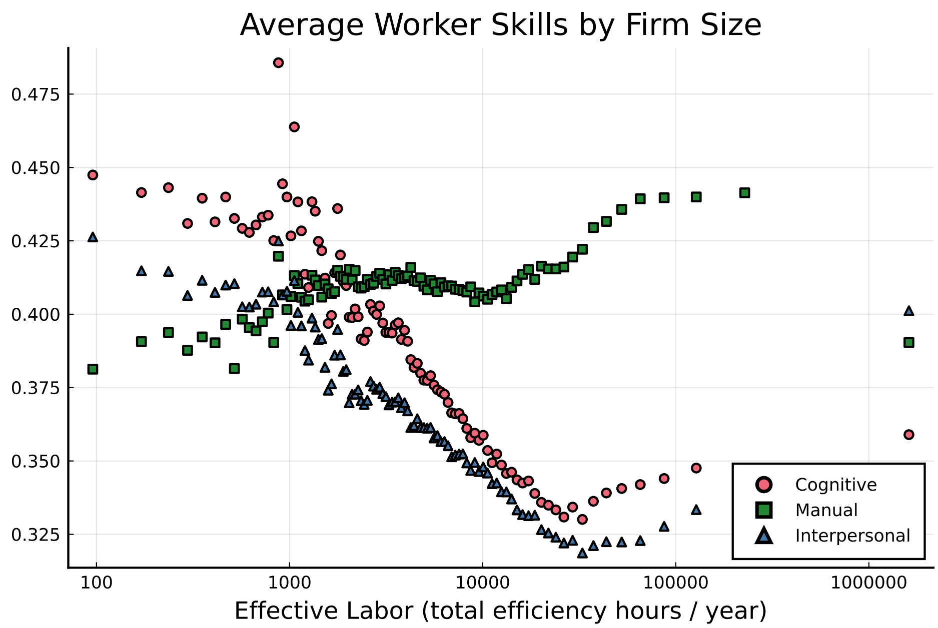


Figure 2: Average worker skills by firm size, in 2000. Firms are ranked by the total number of efficiency units of labor that they hire in each year, and we compute, for each percentile bin, the average quantity of labor hired in that bin, and the average value of cognitive, manual, and interpersonal skills of all the workers hired by firms in this bin, weighted by efficiency hours. Firm size is on log-scale.

data.

Within each percentile bin of firm size, I calculate the mean skill level of the workers hired, weighted by their effective labor supplied, for the cognitive, manual, and interpersonal measures of skill, and plot the results in Figure 2. I find that while firms' choices of the average manual skills of their workforce does not vary much in firm size, both the cognitive and interpersonal skills exhibit a non-monotone shape. That is, while very small and very large firms hire a very highly skilled workforce in the cognitive and interpersonal dimensions, medium sized firms hire, on average, less skilled workers.

To understand the mechanisms at play, I divide the set of firms into "large firms" (50th percentile to 99th percentile) and "small firms" (1st percentile through 49th percentile), and compute the kernel density estimate of the distribution of worker skills in the cognitive and interpersonal dimensions. I plot the resulting estimate of the distribution in fig. 3. Small

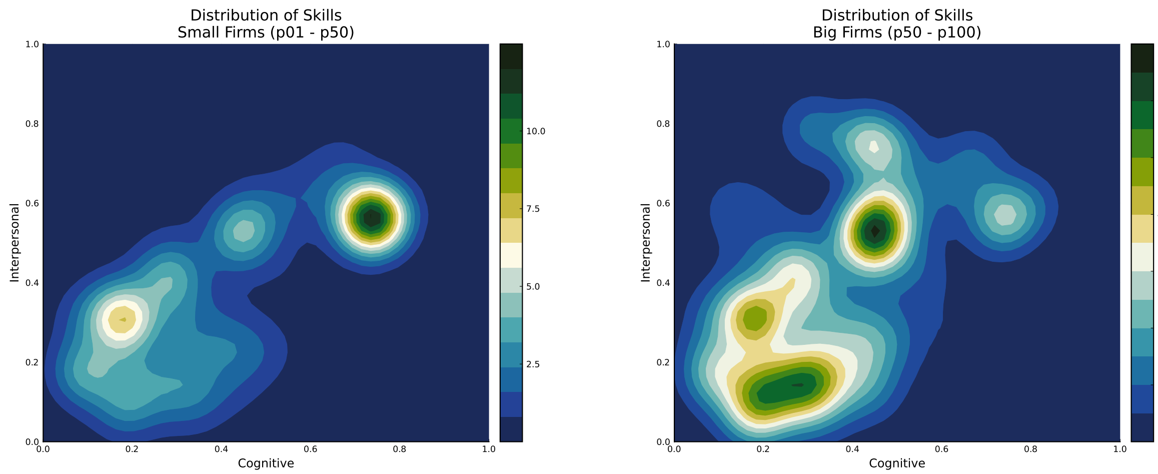


Figure 3: Distribution of worker skills hired, small versus large firms, in 2000. Firms are ranked by the total number of efficiency units of labor that they hire in each year, and divided into large and small firms at the median. For each grouping of firms, I compute the kernel density estimate of the distribution of worker skills, weighted by efficiency hours, in the cognitive and interpersonal dimensions.

firms hire workers with relatively high cognitive and interpersonal skills, with most of the mass concentrated in just one occupation (Sales Engineers), with relatively high cognitive skills (0.73) and interpersonal skills (0.56). In contrast, among larger firms, the single largest occupation hired are secretaries and administrative assistants (cognitive skills of 0.44 and interpersonal skills of 0.51), and the occupations are *much* more dispersed in skill space, with only 7.1% of the effective concentrated at the mode (as opposed to 22.9% in small firms). Much of the mass in the distribution moves to workers with relatively low skills, but who are more dispersed in the skill space, consistent with firms choosing to hire less skilled workers, who are better matched to the tasks that they are assigned to do in production.

I quantify the increase in the dispersion that we observe by calculating, for each firm, the within-firm standard deviation of worker skills, in the cognitive, manual, and interpersonal dimensions. This is a firm-specific measure of how dispersed the workers that they choose to hire are in the skill-space. I compute the average of this measure for each percentile of firms, and plot the results in Figure 4. A large share of the increase in dispersion that we observe is being driven by the *within firm* dispersion of worker skills, which suggests an important role

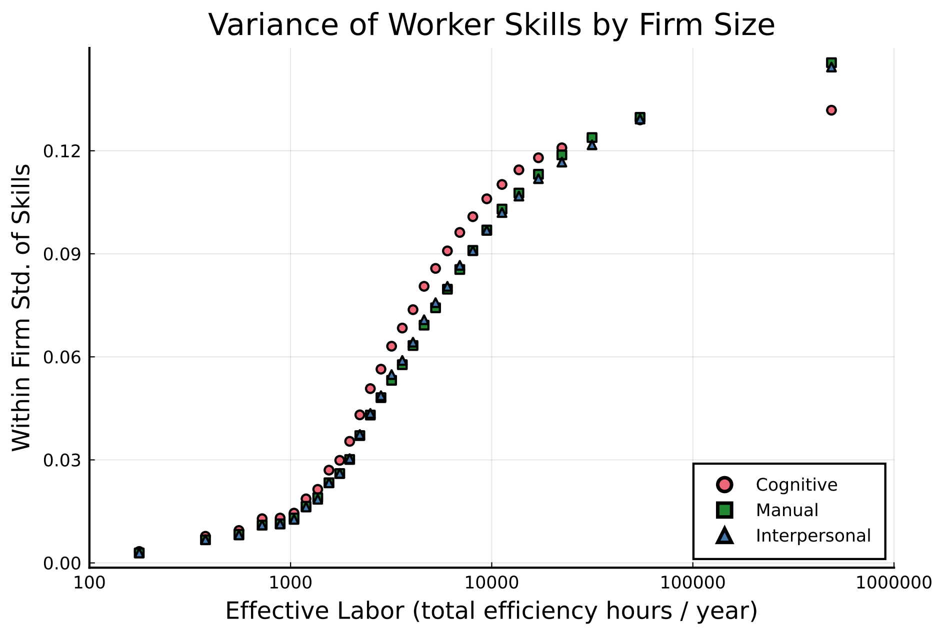


Figure 4: Within-firm dispersion in worker skills, in 2000. Firms are ranked by the total number of efficiency units of labor that they hire in each year, and I compute the within-firm standard deviation of cognitive, manual, and interpersonal skills, weighted by efficiency hours. For each percentile bin, I compute the average quantity of labor hired in that bin, and the mean of the within-firm standard deviations of worker skill. Firm size is on log-scale.

for worker specialization within the firm in understanding how firms choose to hire different types of workers.

To formalize these stylized facts, and to account for industry-specific composition effects, I consider a regression specification that projects the following onto firm size deciles: the log of the number of distinct occupations; the mean skills in each of the cognitive, manual, and interpersonal dimensions; and the within-firm standard deviation of each skill. These are regression-based analogues of Figures 1, 2 and 4. I estimate

$$y_i = \sum_{s=2}^{10} \beta_s D_i^s + \gamma_{d(i)} + \epsilon_i \tag{1}$$

where D_i^s is an indicator variable for whether firm i is in decile s of the firm size distribution (calculated using the firm's total quantity of effective labor hired), and $\gamma_{d(i)}$ is a fixed effect for each 5-digit industry $d(i)$. I report the results from these specifications in Table 1.

In specification (1), I find that firms in the highest decile of the size distribution hire 4.4 times as many distinct occupations on average as firms in the fifth decile, and 6.3 times as many as firms in the lowest decile. The average number of distinct occupations is tightly estimated, with non-overlapping confidence intervals, even when controlling for composition effects by industry.

In specifications (2) - (4), I find that the within-firm dispersion of skills is increasing in each of the three dimensions of skill that I measure. Again, the decile fixed effects are extremely precisely estimated, and are monotonically increasing in each dimension of skill, with non-overlapping confidence intervals. Firms in the highest decile of the size distribution hire a set of workers whose skills are substantially more dispersed than firms in the fifth decile. My estimates imply that the within-firm standard deviation among the largest firms is more than 2.5 times that of the fifth decile firms in all three dimensions of skill. Because these measures of within-firm dispersion in worker skill are weighted by the total quantity of labor hired, this increase in the variance in skills is *not* a mechanical result of having more workers. Workers employed by larger firms genuinely have a more dispersed set of skills than the workers at their smaller counterparts.

Moreover, I find that the average skill level varies substantially by firm size in all three dimensions of skill. In specifications (5) - (7), I find that the average level of skill differs substantially across the deciles of firm size. The differences in average skill are large relative to the within-firm dispersion of skills. For instance, the difference between the average cognitive skills of the workers of workers at the fifth decile and workers at the first decile corresponds to more than half the standard deviation of cognitive skills within fifth decile firms. This difference in average skills corresponds to about 7.6% of the sample average of cognitive skills. Similarly, the difference for firms at the ninth decile corresponds to 13.4% of the sample mean. Crucially, for all three dimensions of skill, we can reject (with p-values below machine precision) the null hypothesis that average skills are constant in firm size (i.e, that all of the coefficients are the same).

I show, in Appendix F that these empirical findings persist even in extremely narrowly defined industries that produce a homogeneous good, such as sugar cane production, coffee production, concrete manufacturing, and the plywood industry. This suggests that my main empirical findings are being driven by an increase in the degree of specialization of workers among larger firms, rather than industry composition effects, or heterogeneity in the goods that are being produced.

To rationalize these findings, and to quantify the role of worker specialization within the firm, in the next section I develop a structural model of how firms choose which occupations to hire, and how to optimally assign tasks to those workers.

	Within Firm Std Skills				Avg Skills		
	log(Occupations)	Cognitive	Manual	Interpersonal	Cognitive	Manual	Interpersonal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
deciles: 2	0.051***	0.006***	0.006***	0.006***	0.003	0.011***	-0.004***
	(0.004)	(4.964e-04)	(4.360e-04)	(4.567e-04)	(0.002)	(0.001)	(0.002)
deciles: 3	0.119***	0.016***	0.014***	0.014***	-0.014***	0.014***	-0.019***
	(0.006)	(7.749e-04)	(9.592e-04)	(8.184e-04)	(0.003)	(0.002)	(0.002)
deciles: 4	0.220***	0.030***	0.026***	0.026***	-0.021***	0.018***	-0.025***
	(0.011)	(0.001)	(0.002)	(0.001)	(0.004)	(0.003)	(0.002)
deciles: 5	0.368***	0.051***	0.042***	0.043***	-0.025***	0.018***	-0.027***
	(0.016)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
deciles: 6	0.521***	0.069***	0.057***	0.058***	-0.030***	0.017***	-0.030***
	(0.019)	(0.003)	(0.003)	(0.002)	(0.003)	(0.005)	(0.003)
deciles: 7	0.706***	0.086***	0.074***	0.074***	-0.035***	0.015**	-0.032***
	(0.020)	(0.003)	(0.003)	(0.002)	(0.004)	(0.007)	(0.005)
deciles: 8	0.927***	0.102***	0.090***	0.089***	-0.038***	0.012	-0.033***
	(0.020)	(0.003)	(0.004)	(0.003)	(0.006)	(0.009)	(0.008)
deciles: 9	1.212***	0.116***	0.106***	0.104***	-0.043***	0.014*	-0.037***
	(0.021)	(0.003)	(0.004)	(0.003)	(0.006)	(0.008)	(0.007)
deciles: 10	1.846***	0.127***	0.123***	0.119***	-0.044***	0.022***	-0.038***
	(0.047)	(0.003)	(0.003)	(0.003)	(0.005)	(0.007)	(0.006)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1330135	1330135	1330135	1330135	1330135	1330135	1330135
R^2	0.632	0.326	0.369	0.388	0.344	0.410	0.352

Table 1: Average skills and within-firm standard deviation of skills are calculated by as the mean/standard deviation of the cognitive, manual, and interpersonal skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired. Industry fixed effects are at the 5-digit industry level, and standard errors are clustered by industry.

4 Model

4.1 Aggregation of Output

Consider a continuum of firms, with a unit mass. Each firm, indexed by $j \in [0, 1]$, produces a differentiated good by hiring labor to complete tasks.

Final output Q is produced by a competitive firm using the output q_j of the continuum of intermediate firms.

$$Q = \left[\int_0^1 q_j^\sigma dj \right]^{\frac{1}{\sigma}} \tag{2}$$

The price index P is given by

$$P = \left[\int_0^1 p_r(q_j)^{\frac{\sigma}{\sigma-1}} dj \right]^{\frac{\sigma-1}{\sigma}} \tag{3}$$

and the inverse demand function for the output of each intermediate firm is given by

$$p_r(q_j) = P \left(\frac{q_j}{Q} \right)^{\sigma-1} \tag{4}$$

Final goods producing firms aggregate the output of the intermediate goods producing firms, using a CES aggregation technology, to produce output to be sold to the household.

4.2 Firms

Each intermediate firm has a set of K discrete tasks $\mathbf{y}_k \in \mathcal{Y} = [0, 1]^d$ which they must complete in order to produce output, and which are defined by their relative difficulty along d different dimensions of skill. The tasks occur in fixed proportions, and are distributed according to a probability mass function $\mathbf{G} \in \Delta^K \subset R^K$. While the firm can scale the total measure of tasks they want to complete up or down by a factor of s, the relative shares of each task must remain fixed. In Figure 5, I show an example distribution $\mathbf{G}$ across six

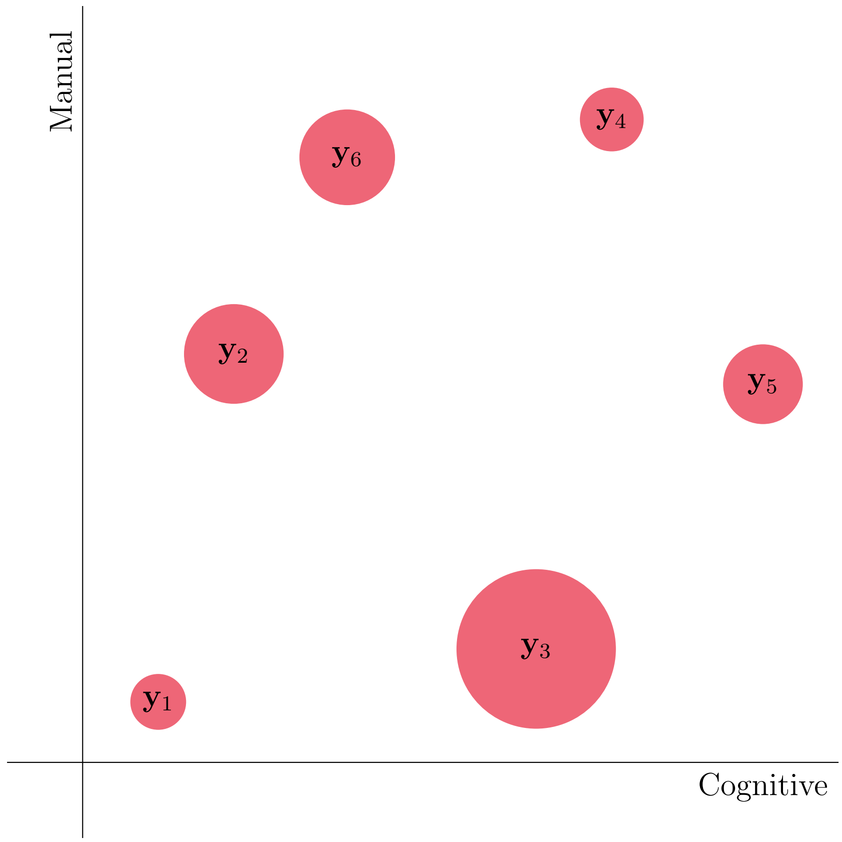


Figure 5: Example: The tasks a firm must complete.

Note: The size of each bubble corresponds to the relative proportion of the tasks

discrete tasks $\mathbf{y}_1, \ldots, \mathbf{y}_6$.

Firms hire workers with a set of skills $\mathbf{x} \in \mathcal{X} = [0,1]^d$ which are a bundle of skills in the same d dimensions. I.e, tasks and skills live in the same space. How "close" an occupation $\mathbf{x}$ is to a task $\mathbf{y}$ tells you how well suited a worker with those skills is to do that task. Workers are endowed with their bundle of skills $\mathbf{x} \in \mathcal{X}$, an idiosyncratic productivity υ, and a unit endowment of time. Their output scales linearly with their idiosyncratic productivity, and so workers can be thought of as supplying υ units of effective labor. For what remains, I will suppress υ for ease of notation, since firms are indifferent between hiring one unit of labor from a worker with $\upsilon = 1$, versus half a unit of labor from a worker with $\upsilon = 2$.

Each firm has a productivity $z_j \in \mathbb{R}$ that are distributed with cdf $F : \mathbb{R} \to \mathbb{R}$. Firms produce output by completing tasks. When a task $\mathbf{y}$ is paired with a worker $\mathbf{x}$, the worker

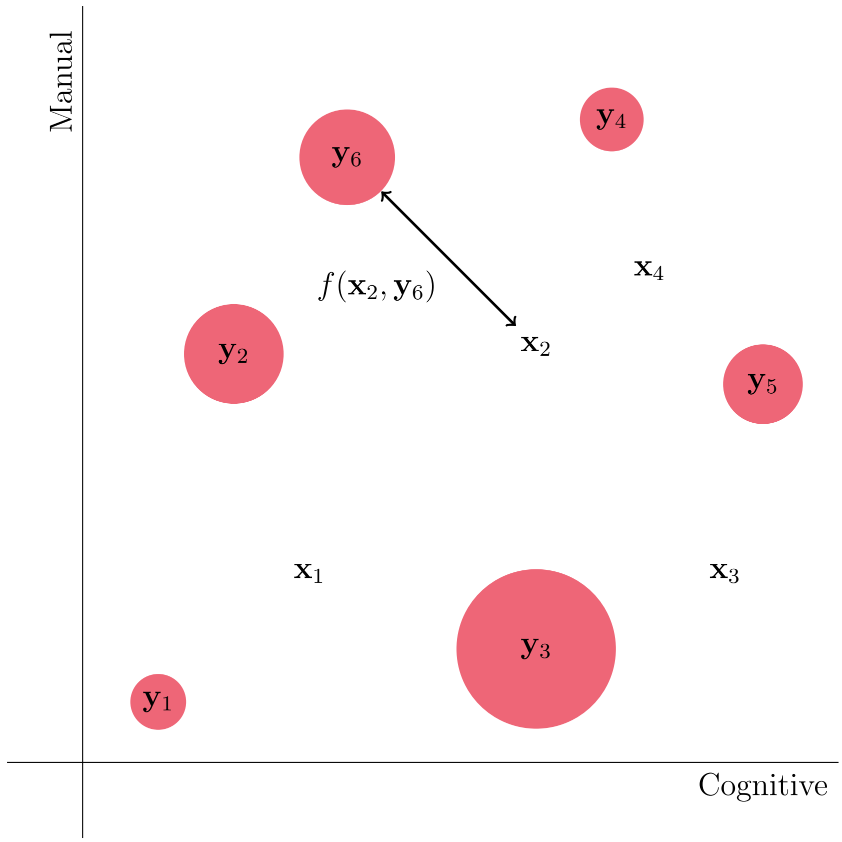


Figure 6: Example Placement of Workers.

Note: When a worker $\mathbf{x}$ is assigned to task $\mathbf{y}$, they produce output of quality $f(\mathbf{x}, \mathbf{y})$

produces $f(\mathbf{x}, \mathbf{y})$ units of output. Firms pay workers a wage $w(\mathbf{x})$, where $w : \mathcal{X} \to \mathbb{R}$ is a competitive wage function that depends on the worker's skill. Firms must pay an organizational cost κ for every discrete worker type that they employ (which captures the additional organizational cost and complexity of managing many different types of workers). As a result, they will choose to hire only a finite number of distinct worker types N in production.

Firms decide how to produce by choosing a *time allocation* $\pi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$. This function takes pairs of workers $\mathbf{x}$ and tasks $\mathbf{y}$, and maps them to the amount of time that workers of type $\mathbf{x}$ will spend working on task $\mathbf{y}$. Since both X and Y are finite sets, with size N and K respectively, it is natural to think about π as a vector in $\mathbb{R}^{NK}$, and I will slightly abuse notation to write $\pi_{nk} = \pi(\mathbf{x}_n, \mathbf{y}_k)$. That is, π_{nk} denotes the amount of time that a worker of type n is assigned to work on task k. If $\pi_{nk} = 0$ that means that task k is not assigned at all to workers of type n. On the other hand, if $\pi_{nk} = \mathbf{G}_k s$, then task k is *exclusively* assigned to workers of type n. The firm's choice of the task assignment π_{nk}, conditional on the choice of $\{\mathbf{x}_n\}$ and $\{\mathbf{L}_n\}$, is an optimal transport problem. Following Villani (2009), I will refer to a task assignment as a *pure assignment solution* if each task is assigned to one and only one worker. That is, if whenever $\pi_{nk} > 0$ for some n and k, that implies $\pi_{n'k} = 0$ for all $n' \neq n$. In Figure 7, I show an example firm's task assignment π_{nk}, in the case where the firm has chosen a pure assignment solution.

The time allocation that the firm chooses must respect two physical constraints. First, the total amount of time that the firm allocates to workers of type n cannot exceed the amount of labor $\mathbf{L}_n$ that the firm has hired from workers of that type. In other words, they must respect the time constraints of the workers that they choose to hire.

$$\sum_{k=1}^{K} \pi_{nk} \leq \mathbf{L}_n \quad n = 1, \ldots, N \tag{5}$$

Second, the total amount of time that the firm allocates to tasks of type n must be *equal* to the amount $G_k s$ of task k which they have chosen to do:

$$\sum_{n=1}^{N} \pi_{nk} = \mathbf{G}_k s \quad k = 1, \ldots, K \tag{6}$$

By requiring that this constraint hold with equality, I insist that every task *must* be done in the given proportions. Although the firm can choose to scale the entire distribution of tasks up or down by a factor s, the relative shares of the tasks must always remain the same. In Figure 8, I show examples of each of these two constraints.

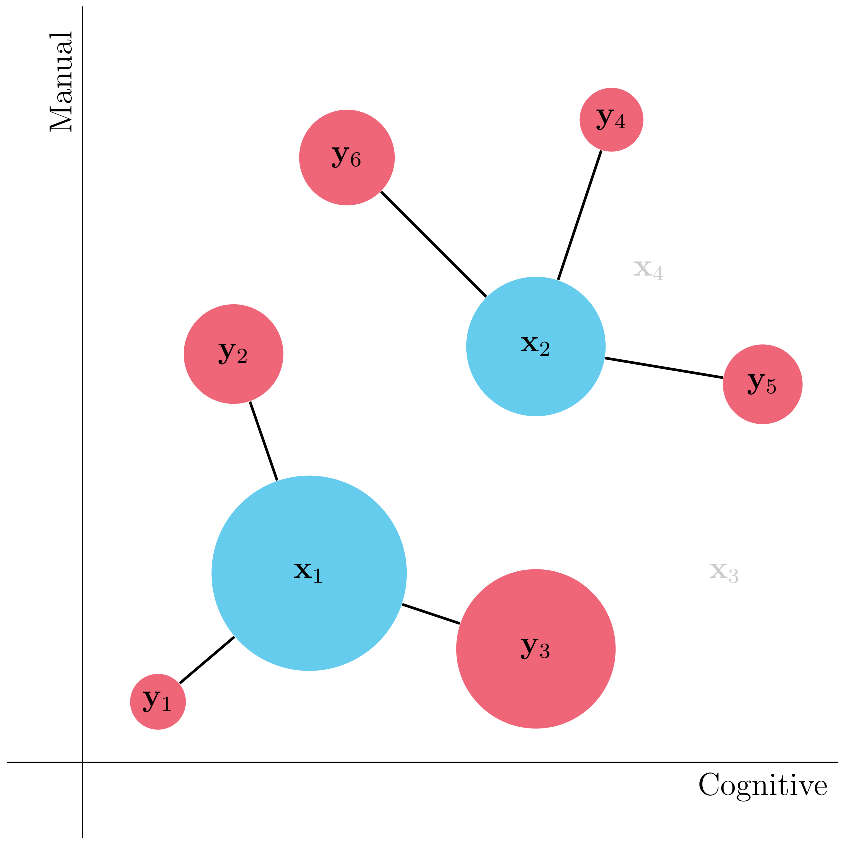
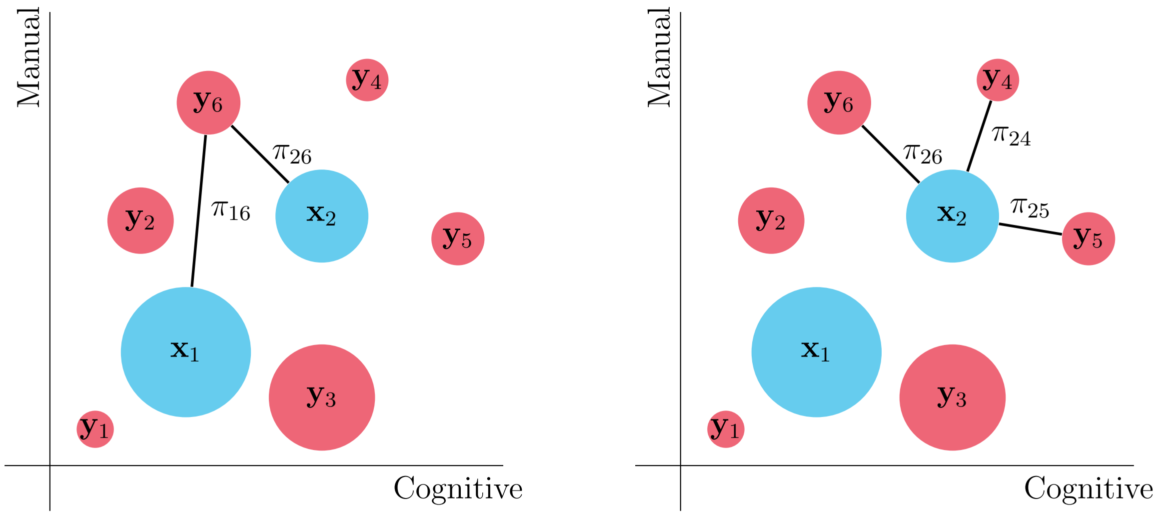


Figure 7: An example worker allocation for a firm.

Note: This firm only chooses to hire $\mathbf{x}_1$ and $\mathbf{x}_2$. Worker types $\mathbf{x}_3$ and $\mathbf{x}_4$ are shown in grey to denote the fact that this firm has chosen not to hire any workers of these types. The size of the circles around $\mathbf{x}_1$ and $\mathbf{x}_2$ denote the quantities of labor $\mathbf{L}_1$ and $\mathbf{L}_2$. In this example, workers of type $\mathbf{x}_1$ are assigned the first three tasks $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$, and workers of type $\mathbf{x}_2$ are assigned the rest. Note: in this example, the firm has chosen a pure assignment, where each task is assigned to one and only one worker. In principle, they could have split a task across several workers.



(a) Each task must be fully assigned.

(b) The workers' time constraints.

Figure 8: Example of the full assignment constraint and the workers' time constraints.

Note: In the panel on the left, we see an example where $\pi_{16} + \pi_{26} = \mathbf{G}_6 s$, reflecting the constraint that each task is always fully assigned. In the panel on the left, we see an example of the workers' time constraint: $\pi_{24} + \pi_{25} + \pi_{26} \leq \mathbf{L}_2$.

Firms aggregate the individual units of output from each worker-task pairing using a CES production technology:

$$q_j = z_j \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \tag{7}$$

where η controls the degree to which output from different worker-task pairings are substitutes or complements.

The firm problem is then to choose:

1. How much output q to produce,
2. How many distinct worker types N to employ in production,
3. The bundle of worker skill vectors $\{\mathbf{x}_n : n = 1, \ldots, N\}$,
4. The measure of each worker type's labor $\mathbf{L}_n$ to hire,
5. The scale of production s,
6. The time allocations π_{nk}

in order to maximize profits.

That is, firms choose their total production, and the number of worker types to satisfy

$$\max_{q} p(q)q - c^N(q, z_j) - \kappa \times N \tag{8}$$

where $c^N(q, z)$ is the cost of producing q units of output. The firm's choice of the number of workers to hire depends on their scale of production. It is key to observe here that the organizational cost $\kappa \times N$ does *not* scale with the the firms output q; it functions as a fixed cost for producing at that organizational scale. A firm with relatively low productivity will choose a lower level of *organizational complexity*, whereas a firm with a higher productivity may be able to support the additional fixed costs of hiring more types of workers, in order to achieve the commensurate gains in worker productivity (and therefore lower per-unit costs of production) due to specialization.

Firms then choose $\{\mathbf{x}_n : n = 1, \ldots, N\}$, $\mathbf{L}$, s, and π to minimize total costs subject to an output constraint, and eqs. (5) and (6)

$$
\begin{aligned}
c^N(q,z) = \min_{\mathbf{x}_n, \mathbf{L}_n, \pi, s} \quad & \sum_{n=1}^{N} w_n \mathbf{L}_n && \text{Total costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} = \mathbf{L}_n \quad \forall n && \text{Every worker is fully utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1}{\eta}} \geq q && \text{Output constraint}
\end{aligned}
\tag{9}
$$

One convenient feature of this formulation of the firm's optimal assignment problem is that all of the choice variables enter linearly other than the choice of the skill bundles $\{\mathbf{x}_n : n = 1, \ldots, N\}$. As a result, we can split this into a two stage budgeting problem, where the firms first choose the bundle of worker skills, and then choose the amount of each worker to hire, the scale of production s, and the time allocation π. This second stage is a linear program, which is extremely tractable both analytically and computationally.

4.3 Properties of the Optimal Assignment

In addition to being a linear program, the firm's optimal assignment problem has a number of desirable properties which I will exploit in the identification and estimation of the model. First, we can note that c^N is linear in $(q/z)^{\eta}$. This is easy to see after noting that since both $z > 0$ and $\eta > 0$, we can always rewrite the firm's output constraint in terms of $(q/z)^{\eta}$:

$$
\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \geq \left(\frac{q}{z} \right)^{\eta} \tag{10}
$$

Since the objective and all of the other constraints are linear in the firm's choice variables, we can factor out $(q/z)^\eta$ from the problem. In particular, this means we can always rewrite the firm's total costs as

$$c^N(q,z) = \left(\frac{q}{z}\right)^\eta c_N^\star \tag{11}$$

for some constant $c_N^\star := c^N(1,1)$. I formalize these claims and prove them in Lemma 1 (see Appendix A).

Furthermore, I show that the firm's choice of the task assignment will always be a *pure assignment solution*. That is, except for knife's edge conditions, the firm will only choose to assign a given task to a single worker. In other words, if $\pi_{nk} > 0$, then $\pi_{n'k} = 0$ for all $n' \neq n$. Intuitively, this is derived from the fact that for a given set of wages and a given production function f, and for each task k, the firm can strictly rank workers by the firm surplus that they generate when assigned to that task. For the firm's task assignment to be optimal, they will always assign task k exclusively to the worker who generates the greatest surplus. I formalize and prove this claim in Proposition 1 (see also Appendix A).

4.4 Identification

In this section, I argue that the key object of inference, the distribution of tasks $\mathbf{G}$, can be identified semi-parametrically using the first order conditions of the firms' problem. Here, I will give an informal description of the identification argument, and state the main result (whose proof can be found in Appendix B).

A necessary condition for optimality that is implied by eq. (9) is that

$$\underbrace{\frac{w(\mathbf{x}_n)\mathbf{L}_n}{\sum_{i=1}^N w(\mathbf{x}_i)\mathbf{L}_i}}_{\text{Worker } n\text{'s share of the wage bill}} = \underbrace{\sum_{k=1}^K f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk}^\star}_{\text{Worker } n\text{'s share of output produced}} + \epsilon_n \tag{12}$$

where ϵ_n is a residual term. That is, each worker's share of firm costs must equal their share of total output produced. I show the derivation for this FOC in Proposition 2, and show that the residual term ϵ_n has mean zero by construction. The key intuition for how we can use this to identify the distribution of tasks is that $\pi^\star_{nk}$ is intimately tied to the distribution $\mathbf{G}$ through the constraints on its marginals. Each time we observe a worker in a different production configuration (i.e., with a different set of co-workers), it provides additional information about what the underlying distribution of tasks must have been.

Each additional firm configuration we observe is guaranteed to provide *at least* one additional degree of freedom to estimate the task distribution, and in practice provides much more additional information than that.

Theorem 1. *Suppose we observe firms that hire up to N occupations in the data, and we observe data on their skills and wages. Suppose further that the function f is known, and that f distinguishes workers almost everywhere. If the number of tasks is K, and $K \leq N$ then the distribution of tasks $\mathbf{G}$ is identified within a neighborhood of the solution.*

For the proof, see Appendix B. This theorem provides a relatively weak guarantee that, for a given production function f, the choices of the largest firms in the economy will pin down a discrete task distribution with as many tasks as the number of workers that they hire. However, this is for the worst possible scenario, where the additional occupation in a firm of size k is assigned a strict subset of the tasks assigned to just a single occupation in the firm of size $N-1$, for all N. (This means that an additional occupation *only* provides information about the distribution of tasks in the first firm size at which they are observed). In practice, this is not an issue, and firms rarely nest the assignment sets of workers (at the parameters I estimate), so I exploit the fact that workers are observed being assigned different tasks at different firms.

To address this, I use a parametric functional form for the distribution of tasks with fewer degrees of freedom than the number of occupations hired in the largest firm, so that I am guaranteed that the parameters are identified. I use the remaining first-order conditions from the smaller firms to identify the parameters of the production function f and the elasticity of substitution parameter η.

5 Estimation

There are four crucial objects in the model that must be estimated: the distribution of tasks $G(\mathbf{y})$, the worker task production function $f(\mathbf{x}, \mathbf{y})$, the distribution of firm productivities $F(z)$, and the fixed costs of hiring an additional occupation κ. To estimate them, I proceed in three stages:

1. I assume a parametric functional form for both $G(\mathbf{y})$ and $f(\mathbf{x}, \mathbf{y})$, and estimate these parameters and the elasticity of substitution η using nonlinear GMM on the moment conditions implied by the firm's problem eq. (9).

2. Given these estimates from stage 1, I recover firm level estimates of both output q and productivity z by choosing the ratio q/z to match the firm's observed wage bill, and backing out q and z from the first order conditions for the firm's choice of q in eq. (8)

3. I then estimate κ using simulated method of moments to match the observed relationship in the data between the firm's wage bill and the total number of occupations hired.

A key identifying assumption here is that the set of tasks being done by firms is constant within an industry. That is, large firms are doing the same set of tasks as small firms. I show, in Appendix G that the overall qualitative results presented in this section, and in Section 6, are robust to a more narrow definition of industry size, in an industry (Leather working) which specializes in a homogeneous good.

5.1 Stage 1: Task distribution and production function

In order to proceed, I fix a set of tasks as the cartesian product of uniformly spaced grids in each dimension of skill:

$$\mathcal{X} = \mathcal{X}_C \times \mathcal{X}_M \times \mathcal{X}_I$$
$$\mathcal{X}_d := \{i/n_d \mid i = 1, \ldots, n_d\} \qquad \text{for } d \in \{C, M, I\}$$

where n_C, n_M and n_I are the number of of grid points in each dimension of skill. I set $n_C = n_M = n_I = 8$, which is a tradeoff between the accuracy of the approximation to a continuous distribution of tasks and the computational burden of solving the firm's optimal assignment problem.

I assume that the distribution of tasks $G(\mathbf{y})$ has marginal distributions that are parameterized by a Beta distribution, with shape parameters α_d and β_d, for each dimension of skill $d = C, M, I$. Beta distributions are a relatively flexible family of distributions that have bounded support on the unit interval. I parameterize the full distribution G using a Gumbel Copula, with rank parameter θ. Here, θ governs the strength of the correlation among the ranks in each dimension.

For a given distribution $G(x)$, I discretize it over the set $\mathcal{X}$ by finding the vector $\mathbf{G} \in \mathbb{R}^K$ such that

$$G(\mathbf{x}_k) = \sum_{r=1}^{K} \mathbb{1}(\mathbf{x}_r \leq \mathbf{x}_k)\mathbf{G}_k \tag{13}$$

In other words, finding the vector of probability masses $\mathbf{G}$ such that the "empirical cdf" of the discretization coincides with G at each of the grid points.

To parameterize the worker task production function, I assume that

$$f(\mathbf{x}, \mathbf{y}) = h\left(\mathbf{x}' A \mathbf{y} - (\mathbf{x} - \mathbf{y})' B (\mathbf{x} - \mathbf{y})\right)$$
$$h(x) = \frac{1}{1 + \exp(-x)}$$

The inverse logit link function h bounds the worker-task output to lie between zero and one. There are two main terms within the production function here: the $\mathbf{x}'A\mathbf{y}$ captures a notion of absolute advantage in the production function. The matrix A encodes how worker skill $\mathbf{x}$ interacts with the skill content of the tasks $\mathbf{y}$. The matrix B, which I restrict to be positive definite, defines a distance metric between worker and tasks. It captures the *comparative advantage* of workers. The quality of a worker's output when assigned to a task is, all else being equal, decreasing in the distance between the worker and the tasks. B controls which dimensions of mismatch are the most important for determining output.

For ease of notation, let $\Theta := (\alpha, \beta, A, B, \theta, \eta)$ denote the collection of all the parameters we are estimating. I estimate the model on the subset of workers and firms in the manufacturing sector.

To actually estimate the model, I calculate, for every firm j, and each occupation n, the residual of the firm's moment condition eq. (12):

$$R^{j,n}(\Theta) = \frac{w_n \mathbf{L}_{j,n}}{\sum_{s=1}^{K} w_s \mathbf{L}_{j,s}} - \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k; \Theta)^{\eta} \pi^{\star}_{j,n,k}(\mathbf{L}_j, \Theta) \tag{14}$$

where $\pi^{\star}_{j,n,k}$ is the solution to eq. (9) under the additional restriction that $\mathbf{L} = \mathbf{L}_j$ (the firms hire the same quantity of effective labor that we observe in the data).[7] To solve the firm's optimal assignment problem, under these constraints, I treat the firm's constrained problem as an optimal transport problem and approximate its solution numerically by solving an entropic regularization of the original problem using Sinkhorn iterations.[8] Rather than

[7]Because the firm's cost function is linear in $(q/z)^{\eta}$, I know that all firms that choose the same number of workers K will choose the same set of which occupations to hire, and will choose a time allocation for their workers which are invariant up to a proportional scaling factor. I exploit this property to estimate θ using moment conditions which are calculated not at the firm level, but aggregated across all firms who choose the same set of occupations in the data. This yields considerable savings in the computational burden of estimating the model. I write the moment conditions at the firm-level throughout for ease of notation.

[8]See Peyré and Cuturi (2019), Chapter 4 for a detailed numerical description of the algorithm, and a proof that for sufficiently small values of the regularization parameter, the solution to this relaxation of the original problem approximates the true solution arbitrarily well. I choose a regularization parameter $\epsilon = 0.01$, which is in line with the values commonly used in the numerical OT literature. This regularization closely approximates the exact solution to the linear program, while converging nearly 3 orders of magnitude

	α	β
Cognitive	3.517	1.595
Manual	3.579	1.820
Interpersonal	1.015	1.148

Table 2: Parameter Estimates for Marginal Distributions of $G(\mathbf{x})$

	Absolute Advantage			Comparative Advantage		
	Cognitive	Manual	Interpersonal	Cognitive	Manual	Interpersonal
Cognitive	1.575	-1.264	-2.161	3.499	0.687	-0.431
Manual	-8.611	2.287	4.617	0.687	0.267	0.008
Interpersonal	9.961	-0.701	-2.686	-0.431	0.008	0.123

Table 3: Parameter Estimates for the production function parameters A and B

targeting eq. (14) directly, I define my moment targets in percentage deviations from the wage bill shares $\widehat{R}^{j,n}(\Theta)$. I then choose the parameters Θ in order to minimize the nonlinear GMM objective:

$$\min_{\Theta} \sum_{j=1}^{J} \sum_{n=1}^{N} \widehat{R}^{j,n}(\Theta)^2 \omega_{j,n} \tag{15}$$

where $\omega_{j,n}$ are weights that are proportional to the total quantity of effective labor employed (so that more weight is put on the FOCs that correspond to a larger employment share of the population). Note that I define both $\widehat{R}^{j,n}(\Theta) = 0$ and $\omega_{j,n} = 0$ if firm j does not employ any workers of occupation n. I report the point estimates in Tables 2 and 3. I plot the implied distribution of tasks in Figure 9. My estimates imply that the interpersonal skill requirements of tasks are much more widely dispersed through the skill space than either manual or cognitive skills. In both the cognitive and the manual dimensions, the bulk of the distribution of tasks is concentrated among tasks with low skill requirements, although there are some higher skill tasks which must be completed. The interpersonal skill requirements of tasks are relatively high, and fairly evenly dispersed through the task space.

faster than exact methods in my application.

5.2 Stage 2: Distribution of firm productivity

Once I have estimates of the task distribution, and the parameters of the firm's production function, I recover estimates of firm productivity and output using the firm's total wage bill, and the first order condition governing their optimal choice of output q. I exogenously set the parameter σ (which governs the elasticity of substitution between firms) to 0.85, which is in the middle of the range of plausible estimates in Atkeson and Burstein (2008).

For each firm j, I know their choice of effective labor hired $\mathbf{L}_{j,n}$ for each occupation n. I also know that their cost function takes the form

$$c^{N_j}(q,z) = \left(\frac{q_j}{z_j}\right)^{\eta} c_j^{\star}$$

where $c_j^{\star}$ is the cost of producing z_j units of output (i.e, when $q_j/z_j = 1$.) Using my estimates from Stage 1, I can calculate the model's implied $c_j^{\star}$, and then I back out the ratio q_j/z_j from the ratio of the firm's wage bill to their "unit" costs:

$$\overline{q}_j := \left(\frac{q_j}{z_j}\right) = \left(\frac{\sum_{n=1}^{N} w_n \mathbf{L}_{j,n}}{c_j^{\star}}\right)^{\frac{1}{\eta}} \tag{16}$$

I show in appendix C that given $\overline{q}_j$, both q and z are separately pinned down by the firm's optimality condition in their choice of q, subject to a normalization of the aggregate price index P to 1.[9] They are given in closed form by:

$$\begin{aligned} q_j &= \left(\frac{\eta \overline{c}_j^{\star} \overline{q}_j}{\alpha\sigma}\right)^{\frac{1}{\sigma}} \\ z_j &= \frac{1}{\overline{q}_j}\left(\frac{\eta \overline{c}_j^{\star} \overline{q}_j}{\alpha\sigma}\right)^{\frac{1}{\sigma}} \end{aligned} \tag{17}$$

where $\alpha = P/Q^{\sigma-1}$. My estimates inherit their log-normal shape from the log-normal dis-

[9] This normalization, while somewhat nonstandard, follows from the fact that I have already specified a unit of account for the wages (multiples of the Brazilian monthly minimum wage), and have *not* pinned down the scale of the distribution of firm productivities.

	Parameter Values
η	0.922
θ	6.699
κ	20.925

Table 4: Parameter Estimates for remaining parameters

tribution of firm wage bills, although the distribution of model-implied firm productivities appears to be left-skewed.

5.3 Stage 3: Fixed costs

To estimate the fixed costs, I solve the firm's full assignment problem eq. (9) for each firm size $N = 1, \ldots, N_{\text{Max}}$. With these costs $\overline{c}_N^\star$ in hand, for any given value of κ, it is easy to simulate the firms' choice of how many occupations to hire, since for any given z, and choice of N, we know their optimal choice of q in closed form:

$$q^\star(z, k) = \left(\frac{\alpha \sigma z^\eta}{\eta \overline{c}_K^\star} \right)^{\frac{1}{\eta - \sigma}} \tag{18}$$

Firms choose N to maximize profits as in eq. (8). I then estimate the fixed costs κ using simulated method of moments, to match the coefficients of the regression of

$$\log(N_j) = \beta_0 + \beta_1 \log(C_j) + \epsilon_j \tag{19}$$

where $C_j = \sum_{n=1}^{N} w_n \mathbf{L}_{j,n}$ is the firm's total wage bill. My estimate of κ, reported in table 4, is denominated in multiples of the Brazilian monthly minimum wage. This value of the fixed cost corresponds to the annual earnings of about 13 workers hired full time at the minimum wage, which suggests that the organizational costs of hiring additional occupations are a considerable burden to firms.

5.4 Decomposing Firm TFP

In this section, I show one of the main results of the estimation procedure: that a substantial share of the overall variance in firm TFP is actually due to the endogenous specialization channel. Consider the firm's production function in eq. (7). If we rewrite the time allocation π_{nk} as $p_{nk}L$ where L is the total quantity of labor hired by the firm, and p_{nk} is the share of total labor allocated to workers of type n working on task k, then we see that we can rewrite firm j's total output as

$$q_j = z_j \underbrace{\left(\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta p_{nk} \right)^{\frac{1}{\eta}}}_{\rho_j} L^{\frac{1}{\eta}} \tag{20}$$

Denote the average output quality term by ρ_j. In other words, we can write firm output as $q_j = z_j \rho_j L^{\frac{1}{\eta}}$. What we observe as firm TFP is actually a composite of the exogenous firm productivity z_j and the endogenous productivity of the workers hired by the firm, ρ_j, which depends on the firm's choice of which workers to hire, and which tasks to assign them in production. We will call the "observed" TFP $\hat{z}_j := z_j \rho_j$. This representation of firm TFP suggests a natural decomposition of the log variance:

$$\mathrm{Var}(\log \hat{z}_j) = \underbrace{\mathrm{Var}(\log z_j)}_{\text{Exogenous Firm Productivity}} + \underbrace{\mathrm{Var}(\log \rho_j)}_{\text{Worker Productivity}} + \underbrace{2\mathrm{Cov}\left(\log z_j, \log \rho_j\right)}_{\text{Endogenous Specialization}} \tag{21}$$

The first term here represents variation in the true exogenous firm productivity across firms. The second term, $\mathrm{Var}(\log \rho_j)$ captures the fact that different mixes of worker types in production are more or less productive. The final term in the decomposition, $\mathrm{Cov}(\log z_j, \log \rho_j)$, captures the fact that more productive firms *endogenously choose a more productive mix of workers* to use in production.

Using the estimates from stages 1 and 2, I have estimates of both q_j and z_j at the firm level. And ρ_j can be estimated by solving the firm's problem of how to choose the optimal

	Comp	Share
$\text{Var}(\log(z_j \rho_j))$	0.128	—
$\text{Var}(\log(z_j))$	0.033	25.656
$\text{Var}(\log(\rho_j))$	0.048	37.516
$2\ \text{Cov}(\log(z_j), \log(\rho_j))$	0.047	36.828

Table 5: Results of the variance decomposition of firm TFP in eq. (21)

time allocation, subject to the additional constraints that the share of workers be fixed at the levels we observe in the data, firm by firm. With these estimates, I calculate the results of the variance decomposition in eq. (21) and report the results in Table 5.

I find that variation in the exogenous firm productivity, which could also capture differences across firms in their capital stocks, accounts for only 24.2% of the variation in overall firm TFP. A larger share of the variation is due to variation in worker productivity across the firms. However, 36.1% of the overall variation in firm TFP is due to the the covariance terms alone. That is, more productive firms endogenously choose to hire a more productive mix of workers in production, and this endogenous specialization channel accounts for more than a third of the observed variation in firm productivity.

Crucially, this endogenous specialization is *not policy invariant.* Any government policies that distort the firms choice of whether or not to hire more specialized workers in production (for instance, a tax on value added) can have large and counterintuitive implications for aggregate productivity. The key observation is that even though firms in this economy have market power, and decreasing returns to scale for a given number of occupations, the fixed costs of hiring an occupation actually give firms a source of increasing returns to scale, which is captured in this variance decomposition by the correlation between firm productivity z_j and the endogenously chosen worker productivity ρ_j. In the next section, I explore the implications of this endogenous specialization mechanism under various counterfactual scenarios.

6 Counterfactuals

To quantify the contribution of endogenous worker specialization within the firm to aggregate productivity and output, I will consider two counterfactual exercises. First, I consider reducing the fixed costs κ faced by the firms to zero. This means that all firms in the economy will have costless access to the most specialized production technology, where they can choose to hire any occupations they like. This counterfactual exercise gives an upper bound on the gains from further worker specialization in the economy.

Second, to quantify the productivity gains from the specialization already occurring in the economy, I consider a scenario where κ is set sufficiently large that no firm chooses to hire more than a single occupation. This minimally specialized baseline economy, serves as a lower bound on what productivity and output could be, if firms were not able to divide their tasks among different occupations.

In order to have an apples to apples comparison, I solve the full model at the baseline parameter estimates. For the counterfactual economies, I need to know how wages will adjust to changes in firm demand for labor. For now, I assume a very simple model of labor supply: each worker is endowed with L units of labor, and and idiosyncratic productivity ν. Workers maximize:

$$\max_{n \in 1,\ldots,N} \log(w_n \nu L) - c_n \tag{22}$$

where c_n is a disamenity cost of working in a particular occupation. An equilibrium is a set of wages and quantities such that workers are indifferent between choosing all of the occupations (which are chosen in equilibrium) and the total quantity of labor demanded, integrating across all the occupations, equals the total supply. Because workers are indifferent between working in each occupation, we know that for any occupations n and n' that are both chosen, it must be the case that

$$\log(w_n) - \log(w_{n'}) = c_n - c_{n'} \tag{23}$$

and therefore the relative wages $w_n/w_{n'}$ are pinned down by the difference in occupation-specific disamenities. In equilibrium, all wages adjust by a constant scaling factor λ so that the *total* quantity of effective labor in the economy remains constant.[10]

I solve for the counterfactual economies, and present the results in Table 6. In the first counterfactual, setting κ to zero and giving all firms costless access to the most specialized production technology, I find relatively modest gains to increasing the degree of firm specialization in the economy. Sector GDP rises by about 1.3%, and consumption by slightly more. This reflects the fact that much of the economy's output is already being produced by relatively large firms in the baseline. These firms are *already* operating at a fairly large scale, and do not change their labor demand or output by much in response to a reduction in the cost of hiring more occupations – they have already hired most of the occupations that they need. The average skills demanded in the economy almost do not change at all.

In contrast, there are *much* larger changes when I restrict access to more specialized ways of organizing production (by setting κ to be sufficiently large). We can think of these counterfactual estimates as telling us what the contribution of worker specialization is to aggregate productivity. If I shut down firm's ability to hire specialized workers to complete their tasks, and require that all of their tasks be completed by a single worker type, I find that output falls by 9.5%, and consumption by 11.2%. These estimates serve as bounds on how large the potential output gains/losses can be under various policy scenarios that distort the firm's specialization margin.

[10] I am currently working on relaxing this assumption, and allowing for imperfect substitution between different occupations on the worker side. However, this theory of labor supply may be appropriate for considering long-run adjustments where the cost of acquiring the skills necessary for an occupation corresponds to the cost and disamentity value of going to school to acquire those skills.

In order to relax this, we need the relative wages between occupations to adjust to clear each occupation's labor market separately in general equilibrium. In general, this is a very difficult problem to solve, as it requires repeatedly re-solving the firm's optimal labor choice (a computationally burdensome mixed-integer linear program).

	Baseline	$\kappa = 0$	$\kappa = 2 \times \hat{\kappa}$	$\kappa =$ Large
% Δ Consumption		1.417	-0.727	-11.284
% Δ Wage		0.065	-0.046	-1.009
% Δ Output		1.203	-0.619	-9.676
Cognitive	0.374	0.369	0.375	0.370
Manual	0.431	0.463	0.420	0.321
Interpersonal	0.356	0.337	0.362	0.405

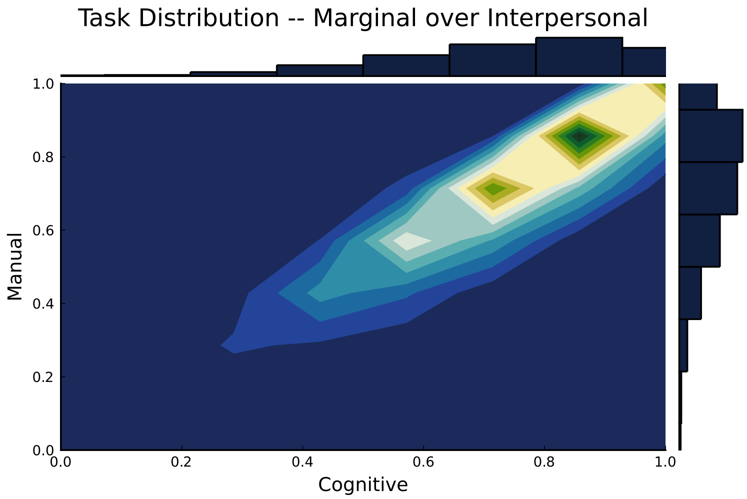
Table 6: Results from Counterfactual Policy Scenarios

7 Conclusion

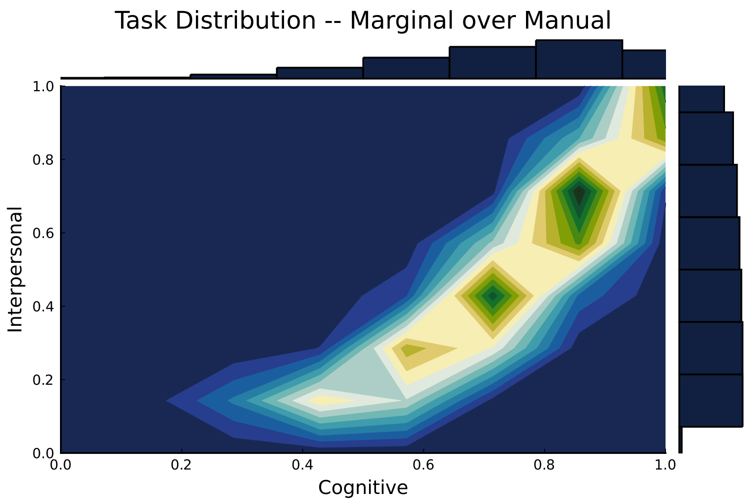
In this paper, I answer the questions: what tasks must be performed to produce a good? Which occupations are well suited to do those tasks, and what are the productivity gains from reorganizing firms to use the optimal mix of occupations to complete these tasks?

My main finding is that firms' endogenous choice of which workers to hire, and how to organize them in production, is of first order importance in understanding both firm specific and aggregate productivity. I find that larger firms make systematically different choices than small firms about which *types* of workers to employ in production. As we move from small to large firms, we see that firms systematically vary the total number of occupations they hire, the average cognitive, manual, and interpersonal skills of those workers, and their degree of dispersion in the skill space.

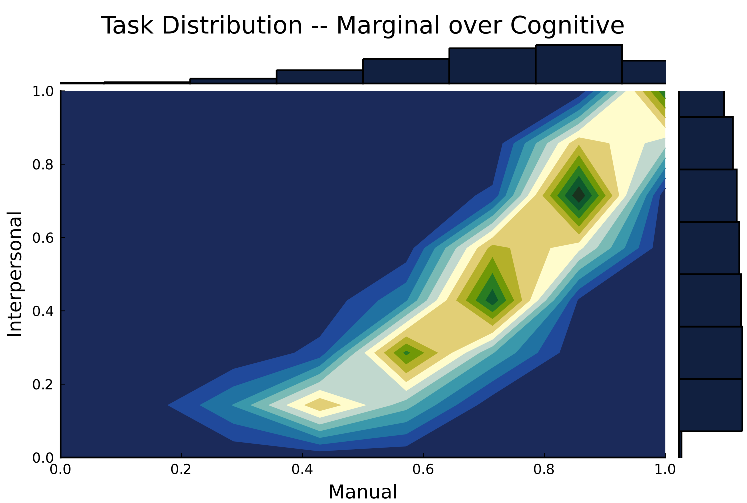
In my quantitative application, I find that 36% of the variation in firm TFP is due to the endogenous choice of firms to hire more specialized workers in production. I find that shutting down this specialization channel leads to output losses of 9.6%, even adjusting wages to hold the total quantity of labor in the economy fixed. Moreover, I find that firms are relatively close to the "optimal" level of specialization. Allowing for *even more* specialization in the economy has gains that are bounded at about 1.3% of sector GDP. These estimates, taken together, suggest that a measure of caution is warranted when considering policies that might distort firms' incentives to specialize workers in production, since there is limited upside, and there are very large potential costs.



(a) Cognitive vs. Manual Tasks



(b) Cognitive vs. Interpersonal Tasks



(c) Manual vs. Interpersonal Tasks

Figure 9: Estimated Distribution of Tasks, Manufacturing, in 2000

References

Acemoglu, D. and D. Autor (2011): "Skills, technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

Ales, L., C. Combemale, E. Fuchs, and K. Whitefoot (2021): "How It's Made: A General Theory of the Labor Implications of Technological Change," Tech. rep.

Armenter, R. and M. Koren (2014): "A balls-and-bins model of trade," *American Economic Review*, 104, 2127–51.

Atkeson, A. and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98, 1998–2031.

Caliendo, L., G. Mion, L. D. Opromolla, and E. Rossi-Hansberg (2020): "Productivity and organization in Portuguese firms," *Journal of Political Economy*, 128, 4211–4257.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015): "The anatomy of French production hierarchies," *Journal of Political Economy*, 123, 809–852.

De Souza, G. (2020): "The Labor Market Consequences of Appropriate Technology," .

Dix-Carneiro, R., P. K. Goldberg, C. Meghir, and G. Ulyssea (2021): "Trade and informality in the presence of labor market frictions and regulations," Tech. rep., National Bureau of Economic Research.

Foster, L., J. Haltiwanger, and C. Syverson (2016): "The slow growth of new plants: Learning about demand?" *Economica*, 83, 91–129.

Friedrich, B. U. (2020): "Trade Shocks, Firm Hierarchies, and Wage Inequality," *The Review of Economics and Statistics*, 1–45.

Galichon, A. (2016): *Optimal transport methods in economics*, Princeton University Press.

Garicano, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of political economy*, 108, 874–904.

Garicano, L. and E. Rossi-Hansberg (2006): "Organization and inequality in a knowledge economy," *The Quarterly journal of economics*, 121, 1383–1435.

Lindenlaub, I. (2017): "Sorting multidimensional types: Theory and application," *The Review of Economic Studies*, 84, 718–789.

Lise, J. and F. Postel-Vinay (2020): "Multidimensional skills, sorting, and human capital accumulation," *American Economic Review*, 110, 2328–76.

Muendler, M.-A., J. Poole, G. Ramey, and T. Wajnberg (2004): "Job Concordances for Brazil: Mapping the Classificaçao Brasileira de Ocupaçoes (CBO) to the International Standard Classification of Occupations (ISCO-88)," *University of California, San Diego, unpublished manuscript*.

Ocampo, S. (2019): "A task-based theory of occupations with multidimensional heterogeneity," JMP.

Peyré, G. and M. Cuturi (2019): "Computational optimal transport: With applications to data science," *Foundations and Trends® in Machine Learning*, 11, 355–607.

Rosen, S. (1978): "Substitution and division of labour," *Economica*, 45, 235–250.

——— (1982): "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, 311–323.

Sanders, C. (2012): "Skill uncertainty, skill accumulation, and occupational choice," *Washington University, St. Louis*.

Tåg, J. (2013): "Production hierarchies in Sweden," *Economics letters*, 121, 210–213.

Villani, C. (2009): *Optimal transport: old and new*, vol. 338, Springer.

A Properties of the Model

To show how we can recover the distribution of tasks, it will help to have established some basic properties of the model.

Lemma 1. *The firm's cost function $c^N(q;z)$ is linear in $(q/z)^\eta$, for all N. In particular, it takes the form $(q/z)^\eta c^\star_N$ for some constant $c^\star_N$ which depends only on the number of occupations hired N.*

Proof. Recall that the firm's cost function c^N is given by

$$\begin{aligned}
c^N(q,z) = \min_{\mathbf{x}_n,\mathbf{L}_n,\pi,s} \quad & \sum_{n=1}^{N} w_n \mathbf{L}_n && \text{Total Costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} = \mathbf{L}_n \quad \forall n && \text{Every worker is fully utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk} \right]^{\frac{1}{\eta}} \geq q && \text{Output Constraint}
\end{aligned} \tag{9}$$

We want to show that for any values of q, z, q' and z', and for any λ, if $(q'/z')^\eta = \lambda(q/z)^\eta$, then $c^N(q',z') = \lambda c^N(q,z)$. So, fix q, z, q', z' and λ. Let us begin by noting that since both q and z are positive, the output constraint can be rewritten as

$$\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk} \geq \left(\frac{q}{z}\right)^\eta$$

Let $\Omega = (\{\mathbf{x}^\star_n\}, \mathbf{L}^\star, \pi^\star, s^\star)$ be a solution to eq. (9) at (q,z).

Claim 1: The tuple $\hat{\Omega} = (\{\hat{\mathbf{x}}_n\}, \hat{\mathbf{L}}, \hat{\pi}, \hat{s}) := (\{\mathbf{x}^\star_n\}, \lambda\mathbf{L}^\star, \lambda\pi^\star, \lambda s^\star)$ is a feasible solution to eq. (9) at (q',z').

To see this, observe that the worker's utilization constraint and the task assignment constraint hold by constraint (since they are just multiplied by λ on both sides). Moreover, we see that

$$\begin{aligned}
\sum_{n=1}^{N}\sum_{k=1}^{K} f(\hat{\mathbf{x}}_n, \mathbf{y}_k)^\eta \hat{\pi}_{nk} &= \sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n^\star, \mathbf{y}_k)^\eta \lambda \pi_{nk}^\star && \text{By definition} \\
&= \lambda \left(\sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n^\star, \mathbf{y}_k)^\eta \pi_{nk}^\star \right) && \text{Factoring out } \lambda \\
&\geq \lambda \left(\frac{q}{z} \right)^\eta && \text{Since } \Omega \text{ is feasible at } (q, z) \\
&= \left(\frac{q'}{z'} \right)^\eta && \text{By assumption}
\end{aligned} \tag{24}$$

Therefore, $\hat{\Omega}$ is feasible at (q', z'). So, by optimality of c^N, we know that

$$\begin{aligned}
c^N(q', z') &\leq \sum_{n=1}^{N} w_n \lambda \mathbf{L}_n^\star && \text{Def of } c^N \\
&\leq \lambda \sum_{n=1}^{N} w_n \mathbf{L}_n^\star && \text{Factoring out } \lambda \\
&\leq \lambda c^N(q, z) && \text{Def of } c^N
\end{aligned} \tag{25}$$

This argument applied in reverse, swapping (q, z) and (q', z'), implies that

$$c^N(q, z) \leq \frac{1}{\lambda} c^N(q', z') \tag{26}$$

which of course means that

$$c^N(q', z') \geq \lambda c^N(q, z) \tag{27}$$

as well. So eq. (25) and eq. (27) together imply that

$$c^N(q', z') = \lambda c^N(q, z) \tag{28}$$

as desired.

To prove the second part of the lemma, let $c_N^\star = c^N(1,1)$. Then we see that for any q and z, if we let $\lambda = (q/z)^\eta$, we see see that $(q/z)^\eta = \lambda(1/1)^\eta$, and therefore $c^N(q,z) = \lambda c^N(1,1) = (q/z)^\eta c_N^\star$. □

This brings us to our first sharp qualitative prediction of the model: after conditioning on the number of different types of workers that the firm chooses to hire, the firm's productivity z *does not* affect the skillset that the firm will choose for those workers, or the relative shares of each worker type that they choose.

Our description of the firm's problem allows for the theoretical possibility that firms may choose to split tasks across workers. That is, for some task k, we only require that the sum of the time each worker spends on task k must equal the total quantity $\mathbf{G}_k s$ which needs to be done:

$$\sum_{n=1}^{N} \pi_{nk} = \mathbf{G}_k s$$

However, there is a special type of time allocation which we will be interested in, where only a single worker is assigned to each task.

Definition 1. For any time allocation π_{nk}, we say that π is a **pure assignment solution** if for all k, if $\pi_{nk} > 0$ then $\pi_{n'k} = 0$ for all $n' \neq n$.

It turns out that no firm will ever find it optimal to split tasks across workers, except on a knife's edge case. In general, for any given task, there is always a worker who is "best suited" to do the task in the sense that the ratio of their output quality to their wage is the highest. If firms were constrained to hire workers in a fixed proportion, they might sometimes be compelled to partially or fully assign a task to the "suboptimal" worker. However, because firms can freely adjust the quantity of labor $\mathbf{L}_n$ hired for each worker type n, they will never find themselves in this situation. They will always hire exactly the correct proportion of workers to line up with the tasks that they are assigned.

We formalize this intuition in the following proposition:

Proposition 1. *For any number of workers N, and any set of workers $\mathcal{X}_N \in \prod_{n=1}^{N} \mathcal{X}$, the firm's optimal time allocation $\pi^\star$ is a pure assignment solution except on a set of measure zero.*

Proof. Consider the first order conditions of the firm's cost minimization problem when $q = 1$ and $z = 1$. Let γ be the multiplier on the output constraint, ρ_n the multiplier on each worker's time constraint, and λ_k the multiplier on each task's time constraint.

We can write the Lagrangian for the problem as follows:

$$\begin{aligned}\mathcal{L} = &\sum_{n=1}^{N} \mathbf{L}_n w_n + \gamma \left(1 - \sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk}\right) \\ &+ \sum_{n=1}^{N} \rho_n \left(\sum_{k=1}^{K} \pi_{nk} - \mathbf{L}_n\right) \\ &+ \sum_{k=1}^{K} \lambda_k \left(\sum_{n=1}^{N} \pi_{nk} - s \times \mathbf{G}_k\right) \\ &+ \sum_{n=1}^{N}\sum_{k=1}^{K} \omega_{nk}\pi_{nk}\end{aligned} \tag{29}$$

with the Kuhn-Tucker complementary slackness conditions $\omega_{nk}\pi_{nk} = 0$ for all n, k. Taking first order conditions with respect to all of the choice variables, we find that optimality of the firm's choices requires

$$[\pi_{nk}] \qquad \gamma f(\mathbf{x}_n, \mathbf{y}_k)^\eta = \rho_n + \lambda_k + \omega_{nk} \qquad \forall n, k \tag{30}$$

$$[\mathbf{L}_n] \qquad w_n = \rho_n \qquad \forall n \tag{31}$$

$$[s] \qquad \sum_{k=1}^{K} \lambda_k \mathbf{G}_k = 0 \tag{32}$$

$$[\rho_n] \qquad \sum_{k=1}^{K} \pi_{nk} = \mathbf{L}_n \qquad \forall n \tag{33}$$

$$[\lambda_k] \qquad \sum_{n=1}^{N} \pi_{nk} = \mathbf{G}_k s \qquad \forall k \tag{34}$$

$$[\gamma] \qquad \sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk} = 1 \tag{35}$$

First, we claim that at the optimal solution, $\gamma = \sum_{n=1}^{N} w_n \mathbf{L}_n$. To see that this must be true, consider that

$$\begin{aligned}
\gamma &= \sum_{nk} \gamma f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk} && \text{Output constraint with } q = 1 \\
&= \sum_{nk} (\lambda_k + \rho_n + \omega_{nk})\, \pi_{nk} && \text{Substituting eqs. (30) and (31)} \\
&= \sum_{nk} (\lambda_k + \rho_n)\, \pi_{nk} && \text{Complementary slackness condition} \\
&= \sum_{n=1}^{N} \rho_n \sum_{k=1}^{K} \pi_{nk} + \sum_{k=1}^{K} \lambda_k \sum_{n=1}^{N} \pi_{nk} && \text{Rearranging the sums} \\
&= \sum_{n=1}^{N} \rho_n \mathbf{L}_n + \sum_{k=1}^{K} \lambda_k \mathbf{G}_k s && \text{Substituting feasibility constraints} \\
&= \sum_{n=1}^{N} \rho_n \mathbf{L}_n && \text{Since } \sum_{k} \lambda_k \mathbf{G}_k = 0 \\
&= \sum_{n=1}^{N} w_n \mathbf{L}_n && \text{Substituting eq. (31)}
\end{aligned} \tag{36}$$

Note that if π is not a pure assignment solution, then there exists distinct n and n'

such that for some k, $\pi_{nk} > 0$ and $\pi_{n'k} > 0$. This implies that $\omega_{nk} = \omega_{n'k} = 0$ by the complementary slackness condition. We see that by substituting eq. (30), we obtain

$$\gamma f(\mathbf{x}_n, \mathbf{y}_k) - w_n = \gamma f(\mathbf{x}_{n'}, \mathbf{y}_k) - w_{n'} \tag{37}$$

which occurs only on a knife's edge case for a given set of wages. □

The next proposition shows that a necessary condition for firm optimality is that for each worker, their share of the wage bill must be exactly equal to their share of output in production, except for a residual which corresponds to the average value to the firm of the tasks the worker is assigned in production.

Proposition 2. *If* $\Omega = (\{\mathbf{x}_n\}, \mathbf{L}, \pi, s)$ *solves eq.* (9) *at* $q = z = 1$ *then*

$$\frac{w_n \mathbf{L}_n}{\sum_{i=1}^{N} w_i \mathbf{L}_i} = \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} + \sum_{k=1}^{K} \frac{\lambda_k}{\gamma} \pi_{nk} \tag{38}$$

for all n.

Proof. Consider the first order condition eq. (30). For each occupation n, multiply eq. (30) by π_{nk}, substitute eq. (31), and sum over all the tasks k to obtain

$$\gamma \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} = w_n \sum_{k=1}^{K} \pi_{nk} + \sum_{k=1}^{K} \lambda_k \pi_{nk} \tag{39}$$

Recall that $\gamma = \sum_{i=1}^{N} w_i \mathbf{L}_i$, and that $\sum_{k=1}^{K} \pi_{nk} = \mathbf{L}_n$. Substituting into eq. (39) and rearranging, we obtain the desired result. □

For what remains, we will need the identifying assumption:

Assumption 1. For every worker n, $\sum_{k=1}^{K} \lambda_k \pi_{nk} = 0$

This requires that for each bundle of worker skills, the firm's average surplus from assigning that worker to a task (calculated over the set of tasks they are assigned in production) is zero. We know that averaging over all workers, this must hold exactly – this is eq. (32).

Now, suppose we see a firm hiring N distinct occupations. Since we know that the optimal assignment π is a pure assignment solution (from proposition 1), under assumption 1 we can rewrite eq. (39) as

$$\frac{w_n \mathbf{L}_n}{\sum_{i=1}^{N} w_i \mathbf{L}_i} = \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \delta_{nk} \mathbf{G}_k s \tag{40}$$

where $\delta_{nk} = 1$ if $\pi_{nk} > 0$ and 0 otherwise. It is easy to see that we can rewrite this as a matrix equation:

$$\mathbf{W^N} = s\mathbf{F}^N \times \mathbf{G} \tag{41}$$

where $\mathbf{W}_n^N := \frac{w_n \mathbf{L}_n}{\sum_{i=1}^{N} w_i \mathbf{L}_i}$ and

$$\mathbf{F}^N := \begin{bmatrix} f(\mathbf{x}_1, \mathbf{y}_1)\delta_{11} & \dots & f(\mathbf{x}_1, \mathbf{y}_k)\delta_{1k} & \dots & f(\mathbf{x}_1, \mathbf{y}_K)\delta_{1K} \\ \vdots & & \vdots & & \vdots \\ f(\mathbf{x}_n, \mathbf{y}_1)\delta_{n1} & \dots & f(\mathbf{x}_n, \mathbf{y}_k)\delta_{nk} & \dots & f(\mathbf{x}_n, \mathbf{y}_K)\delta_{nK} \\ \vdots & & \vdots & & \vdots \\ f(\mathbf{x}_N, \mathbf{y}_1)\delta_{N1} & \dots & f(\mathbf{x}_N, \mathbf{y}_k)\delta_{Nk} & \dots & f(\mathbf{x}_N, \mathbf{y}_K)\delta_{NK} \end{bmatrix}$$

It turns out that for any firm of size N, the matrix $\mathbf{F}^N$ has full row rank.

Proposition 3. *If $f(\mathbf{x}, \mathbf{y}) > 0$ then the matrix $\mathbf{F}^N$ has full row rank.*

Proof. Within a firm, the fact that the optimal time allocation π_{nk} is a pure assignment solution (proposition 1) implies that if $\delta_{nk} = 1$ then $\delta_{n'k} = 0$ for all $n' \neq n$.

Now, let $(\mathbf{F}_n^N)_k := \mathbf{F}_{nk}^N$ denote the nth row of $\mathbf{F}^N$. Suppose there exist weights $\{\phi_n\}_{n=1}^N$

such that

$$\sum_{n=1}^{N} \phi_n \mathbf{F}_n^N = 0$$

Then we know that for each k,

$$\sum_{n=1}^{N} \phi_n f(\mathbf{x}_n, \mathbf{y}_k)\delta_{nk} = 0$$

Since δ_{nk} can only be nonzero for a single $n^\star(k)$, we know that this implies

$$\phi_{n^\star(k)} f(\mathbf{x}_{n^\star(k)}, \mathbf{y}_k) = 0 \implies \phi_{n^\star(k)} = 0$$

Note that the mapping $n^\star$ is surjective (every worker type must be assigned at least one task, otherwise the firm would never have paid the fixed cost to hire workers of that type in the first place). But this means that

$$\phi_n = 0 \quad \forall n$$

Thus, all the rows of $\mathbf{F}^N$ are linearly independent, and $\text{rank}(\mathbf{F}^N) = N$. □

B Proof of Theorem 1

By lemma 1 it is sufficient to consider the case where $q = z = 1$. We know that

$$\begin{aligned} 1 &= \sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \pi_{nk} && \text{eq. (35)} \\ &= \sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} \mathbf{G}_k s && \text{By proposition 1} \\ \implies s &= \left(\sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} \mathbf{G}_k\right)^{-1} \end{aligned} \tag{42}$$

So, let's rewrite eq. (40) substituting in the value of s:

$$\mathbf{W}_n^N \left(\sum_{k=1}^{K} \sum_{i=1}^{N} f(\mathbf{x}_i, \mathbf{y}_k)^\eta \delta_{ik} \mathbf{G}_k \right) = \sum_{k=1}^{K} f(\mathbf{x}_n \mathbf{y}_k)^\eta \delta_{nk} \mathbf{G}_k \tag{43}$$

Rearranging and collecting terms, we find that for each n,

$$0 = \sum_{k=1}^{K} \left(f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} - \mathbf{W}_n^N \sum_{i=1}^{N} f(\mathbf{x}_i, \mathbf{y}_k)^\eta \delta_{ik} \right) \mathbf{G}_k \tag{44}$$

Let $\hat{\mathbf{F}}_{nk}^N := f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} - \mathbf{W}_n^N \sum_{i=1}^{N} f(\mathbf{x}_i, \mathbf{y}_k)^\eta \delta_{ik}$, and let $\hat{\mathbf{F}}_n^N$ denote the nth row. Unlike in proposition 3, we cannot show that $\hat{\mathbf{F}}^N$ has full row rank.[11] However, we can show that by dropping a single row of $\hat{\mathbf{F}}$, and replacing it with the constraint that $\sum_{k=1}^{K} \mathbf{G}_k = 1$, we obtain a system with rank N. We proceed in two steps:

1. First, we show that the first $N-1$ rows of $\hat{\mathbf{F}}$ are linearly independent
2. Second, we show that the vector $(1, 1, \ldots, 1)$ does not lie in the span of the first $N-1$ rows

Step 1. As in the proof of Proposition 3, let $n^\star(k)$ denote the unique worker type n such that $\delta_{nk} > 0$. Without loss of generality, let us suppose that $n^\star(K) = N$. That is, the Nth worker is assigned to the Kth task. We can do this without loss of generality, since we can always reorder the list of workers so that it is true.

Now, suppose that there exist weights $\{\phi_n\}_{n=1}^{N-1}$ such that $\sum_{n=1}^{N-1} \phi_n \hat{\mathbf{F}}_n^N = 0$. We want to show that $\phi_n = 0$ for all $n = 1, \ldots, N-1$.

Consider $k < K$. We see that

$$\sum_{n=1}^{N-1} \phi_n f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} = \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n \sum_{i=1}^{N} f(\mathbf{x}_i, \mathbf{y}_k)^\eta \delta_{ik} \tag{45}$$

[11] In fact, it is easy to see that it does not. Sum eq. (44) across all values of n. The first summand $\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^\eta \delta_{nk} \mathbf{G}_k$ is exactly s. The second summand $\sum_{n=1}^{N} \mathbf{W}_n^N s$ is also exactly s, since $\sum_{n=1}^{N} \mathbf{W}_n^N = 1$ by construction. So, we see that the sum of all the rows of $\hat{\mathbf{F}}$ is identically zero.

Since π is a pure assignment solution, know that only a single δ_{nk} is positive for each k. So we can rewrite this as

$$\begin{aligned} \phi_{n^\star(k)} f(\mathbf{x}_{n^\star(k)}, \mathbf{y}_k)^\eta &= f(\mathbf{x}_{n^\star(k)}, \mathbf{y}_k)^\eta \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n \\ \implies \phi_{n^\star(k)} &= \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n \end{aligned} \tag{46}$$

since f is strictly positive. Since $n^\star$ is a surjection from $\{1, \ldots, K-1\}$ to $\{1, \ldots, N-1\}$, and this is true for all $k < K$, this means that all of the weights ϕ_n are a constant $\overline{\phi}$, which solves $\overline{\phi} = \overline{\phi} \sum_{n=1}^{N-1} \mathbf{W}_n^N$. Therefore,

$$0 = \overline{\phi} \left(1 - \sum_{n=1}^{N-1} \mathbf{W}_n^N \right)$$

But since every wage bill share is strictly positive [12], and we know that $\sum_{n=1}^{N} \mathbf{W}_n^N = 1$, that means that $1 - \sum_{n=1}^{N=1} \mathbf{W}_n^N > 0$. This implies that $\overline{\phi} = 0$, which proves the claim.

Step 2. Now, to show that the vector $(1, 1, \ldots, 1)$ does not lie in the span of $\left\{ \hat{\mathbf{F}}_n^N \right\}_{n=1}^{N-1}$, suppose towards contradiction that it does. That is, suppose there exist weights $\{\phi_n\}_{n=1}^{N-1}$ such that for every k

$$\sum_{n=1}^{N-1} \phi_n \hat{\mathbf{F}}_{nk}^N = 1$$

We can substitute in the definition of $\hat{\mathbf{F}}_{nk}^N$ to obtain

$$\phi_{n^\star(k)} f(\mathbf{x}_{n^\star(k)}, \mathbf{y}_k)^\eta = 1 + f(\mathbf{x}_{n^\star(k)}, \mathbf{y}_k)^\eta \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n \qquad \text{for } k < K \tag{47}$$

$$0 = 1 + f(\mathbf{x}_N, \mathbf{y}_K)^\eta \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n \qquad \text{for } k = K \tag{48}$$

[12] If not, that would imply that the firm did not hire any workers of that type, which means that their choice to pay the fixed cost of hiring that occupation could not have been optimal.

Now, eq. (48) implies that

$$\sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n = \frac{-1}{f(\mathbf{x}_N, \mathbf{y}_K)^\eta}$$

Substituting this into eq. (47) we find that

$$\phi_{n^\star(k)} = \frac{1}{f(\mathbf{x}_{n^\star(k)}, \mathbf{y}_k)^\eta} - \frac{1}{f(\mathbf{x}_N, \mathbf{y}_K)^\eta} \tag{49}$$

Let $k^\star(n)$ denote a task (not necessarily unique) for which $n^\star(k^\star(n)) = n$. We know that such a task exists because every worker is assigned at least one task.

But consider that

$$\begin{aligned}
\frac{-1}{f(\mathbf{x}_N, \mathbf{y}_K)^\eta} &= \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_n \\
&= \sum_{n=1}^{N-1} \mathbf{W}_n^N \phi_{n^\star(k^\star(n))} \\
&= \sum_{n=1}^{N-1} \mathbf{W}_n^N \left(\frac{1}{f(\mathbf{x}_n, \mathbf{y}_{k^\star(n)})^\eta} - \frac{1}{f(\mathbf{x}_N, \mathbf{y}_K)^\eta} \right) \\
&= \sum_{n=1}^{N-1} \mathbf{W}_n^N \left(\frac{1}{f(\mathbf{x}_n, \mathbf{y}_{k^\star(n)})^\eta} \right) - \frac{\sum_{n=1}^{N-1} \mathbf{W}_n^N}{f(\mathbf{x}_N, \mathbf{y}_K)^\eta} \\
\implies \quad \frac{-\mathbf{W}_N^N}{f(\mathbf{x}_N, \mathbf{y}_K)^\eta} &= \sum_{n=1}^{N-1} \left(\frac{\mathbf{W}_n^N}{f(\mathbf{x}_n, \mathbf{y}_{k^\star(n)})^\eta} \right) \\
\implies \quad 0 &= \sum_{n=1}^{N} \left(\frac{\mathbf{W}_n^N}{f(\mathbf{x}_n, \mathbf{y}_{k^\star(n)})^\eta} \right)
\end{aligned} \tag{50}$$

This is a contradiction since f is strictly positive and $\mathbf{W}_n^N > 0$. So, $(1, 1, \ldots, 1)$ does not lie in the span of the first $N-1$ rows of $\hat{F}$.

C Recovering q and z

Suppose that for a given firm, hiring k different types of workers, we observe the skill bundles of their workers $\{\mathbf{y}_j : j = 1, \ldots, k\}$, and the effective labor hired M_j for each type. How can we back out the firm quantities and productivities from what we observe?

Start from the fact that $c^k(q,z) = \left(\frac{q}{z}\right)^\eta \overline{c}^k$, where $\overline{c}^k$ is the total (labor) cost to the firm of producing at $q/z = 1$, which is given by the model.

We observe the firm's actual wage bill $W = \sum_{j=1}^k M_j w(\mathbf{y}_j)$, and so we can back out the firms output to productivity ratio $\overline{q}$ as

$$\overline{q} = \frac{q}{z} = \left(\frac{\sum_{j=1}^k M_j w(\mathbf{y}_j)}{\overline{c}^k}\right)^{\frac{1}{\eta}} \tag{51}$$

The firm faces an inverse demand curve of the form $p(q) = \alpha q^{\sigma-1}$. So, eq. (8) can be rewritten as

$$\max_{q,k} \alpha q^\sigma - \left(\frac{q}{z}\right)^\eta \overline{c}^k - \kappa \times k \tag{52}$$

Taking the first order condition with respect to q, we see that a necessary condition for the firm to be producing optimally is that

$$\alpha\sigma q^{\sigma-1} = \eta \left(\frac{q^{\eta-1}}{z^\eta}\right) \overline{c}^k \tag{53}$$

If we substitute for z using the definition of $\overline{q}$, and isolate q, we obtain q in closed form:

$$\begin{aligned} q &= \left(\frac{\eta \overline{c}^k \overline{q}}{\alpha\sigma}\right)^{\frac{1}{\sigma}} \\ z &= \frac{1}{\overline{q}} \left(\frac{\eta \overline{c}^k \overline{q}}{\alpha\sigma}\right)^{\frac{1}{\sigma}} \end{aligned} \tag{54}$$

So, to obtain q and z from the data we observe on each firm, given a value of α we proceed in three steps:

1. Given the firm's choice of worker skills and shares of effective labor, find the firm's unit costs $\overline{c}^k$ using the solution to the firm's assignment problem eq. (9)

2. Back out the output to productivity ratio $\overline{q}$ using eq. (51)

3. Solve for q and z using eq. (54) and the definition of $\overline{q}$.

Now, recall that α is actually a function of the aggregate price index and quantities. We need to make sure that the implied prices and quantities aggregate up to the correct price indices.

D Extension: Outsourcing

In this section, I show how to extend the main model of this paper to accommodate outsourcing of tasks. Here, I consider the model of Section 4, but relax the requirement that all tasks must be performed by workers within the firm.

In addition to the firms which produce the intermediate output goods, there is an additional representative outsourcing firm. This firm has the technology to intermediate between goods producing firms and workers: for any task $\mathbf{y}$, the outsourcing firm provides labor from a worker with exactly the best skills for the job ($\mathbf{x} = y$) and charges a constant markup μ over the marginal cost of providing the labor. Let $w : \mathcal{X} \to \mathbb{R}$ denote the competitive wage function, which encodes the wage paid to every skill level.

Firms can choose to assign each task either to one of their n workers, or to outsource the task to the outsourcing firm on a spot market. They choose both a time allocation π_{nk} and an outsourcing allocation σ_k such that for every task k, the total quantity of time that the in-house workers are assigned to work on the task and the time outsourced workers are assigned to the task are equal to the time requirements of the task. That is,

$$\sum_{n=1}^{N} \pi_{nk} + \sigma_k = s \times \mathbf{G}_k \quad \forall k \tag{55}$$

The firm problem now becomes:

$$
\begin{aligned}
c^N(q,z) = \min_{\mathbf{x}_n, \mathbf{L}_n, \pi, \sigma, s} \quad & \sum_{n=1}^{N} w(\mathbf{x}_n)\mathbf{L}_n + \sum_{k=1}^{K} \mu w(\mathbf{y}_k)\sigma_k && \text{Total costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} = \mathbf{L}_n \quad \forall n && \text{Every worker is fully utilized} \\
& \sum_{n=1}^{N} \pi_{nk} + \sigma_k = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& z\left[\sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta}\pi_{nk}\right]^{\frac{1}{\eta}} \geq q && \text{Output constraint}
\end{aligned}
\tag{56}
$$

Firms trade off the gains from having outsourced workers who are extremely well suited to the tasks being assigned against the increased costs for those workers that are charged by the outsourcing firm. Firms tend to assign workers to tasks when those tasks are fairly close to those workers' skills, and to outsource tasks which are farther away from their chosen workers. Crucially, by choosing which tasks are performed in-house, firms endogenously form their own boundary.

E Extension: Hicks Neutral Capital

In this section, I show how to extend the model in section 4 to accommodate Hicks-neutral capital, and how to separately recover estimates of the firm's capital stock using the first order conditions of the firm's problem.

Suppose that each firm j can purchase capital k_j by renting it from the household, and that they have a Cobb-Douglass production technology (with capital share β) to combine capital and the quality-aggregated task output from their workers. The firm's production technology (originally given in eq. (7)) is now:

$$
q_j = z_j k_j^{\beta}\left[\sum_{n=1}^{N}\sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta}\pi_{nk}\right]^{\frac{1-\beta}{\eta}} \tag{57}
$$

The firm's cost minimization problem can now be written as:

$$
\begin{aligned}
c^N(q,z,k) = \min_{\mathbf{x}_n, \mathbf{L}_n, \pi, s} \quad & \sum_{n=1}^{N} w_n \mathbf{L}_n && \text{Total labor costs} \\
\text{s.t.} \quad & \sum_{k=1}^{K} \pi_{nk} = \mathbf{L}_n \quad \forall n && \text{Every worker is fully utilized} \\
& \sum_{n=1}^{N} \pi_{nk} = s \times \mathbf{G}_k \quad \forall k && \text{Every task is fully assigned} \\
& zk^{\beta} \left[\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \right]^{\frac{1-\beta}{\eta}} \geq q && \text{Output constraint}
\end{aligned}
\tag{58}
$$

Proposition 4. *The firm's cost function in eq.* (58) $c^N(q,z,k)$ *is linear in* $\left(\frac{q}{zk^{\beta}}\right)^{\frac{\eta}{1-\beta}}$. *In particular it takes the form* $\left(\frac{q}{zk^{\beta}}\right)^{\frac{\eta}{1-\beta}} c_N^{\star}$ *for some constant* $c_N^{\star}$ *which depends only on the number of occupations hired* N*.*

Proof. Observe that we can rewrite the firm's output constraint as

$$
\sum_{n=1}^{N} \sum_{k=1}^{K} f(\mathbf{x}_n, \mathbf{y}_k)^{\eta} \pi_{nk} \geq \left(\frac{q}{zk^{\beta}}\right)^{\frac{\eta}{1-\beta}}
$$

Define $\hat{z} = zk^{\beta}$ and $\hat{\eta} = \frac{\eta}{1-\beta}$. From here, the proof is identical to Lemma 1, using $\hat{z}$ and $\hat{\eta}$ instead of z and η. □

An immediate corollary to this proposition is that the solution to the firm's task allocation problem is identical to the task assignment chosen by the firm without capital. After factoring out the factor of $\left(\frac{q}{zk^{\beta}}\right)^{\frac{\eta}{1-\beta}}$ from the problem with capital, and the factor of $\left(\frac{q}{z}\right)^{\eta}$ from the problem without, the two problems coincide exactly.

In the next proposition, I show that given a value of β, we can simultaneously recover estimates of q_j, z_j, and k_j for each firm by matching the firm's overall wage bill and exploiting the optimality of the firm's choice of output and capital.

Proposition 5. *Let* $\overline{q} = \frac{q}{zk^\beta}$ *and let* r *be the rental price of capital. Then if* $k^\star$ *and* $q^\star$ *are the firm's optimal choice of capital and output, we can recover them in closed form as:*

$$\begin{aligned} k^\star &= \left(\frac{\beta\eta}{1-\beta}\right)\left(\frac{c_N^\star}{r}\right)\overline{q}^{\frac{\eta}{1-\beta}} \\ q^\star &= \left[\left(\frac{\eta/(1-\beta)}{\alpha\sigma}\right) \times c_N^\star \times \overline{q}^{\frac{\eta}{1-\beta}}\right]^{\frac{1}{\sigma}} \end{aligned} \tag{59}$$

Proof. Consider now the firm's profit maximization problem for a given number of worker types N:

$$\max_{q,k} \alpha q^\sigma - \left(\frac{q}{zk^\alpha}\right)^{\frac{\eta}{1-\beta}} c_N^\star - kr \tag{60}$$

If we take the first order condition with respect to k, and substitute the definition of $\overline{q}$ we see that a necessary condition for optimality is that

$$\begin{aligned} r &= \left(\frac{\beta\eta}{1-\beta}\right)\left(\frac{c_N^\star}{k^\star}\right)\overline{q}^{\frac{\eta}{1-\beta}} \\ \implies \quad k^\star &= \left(\frac{\beta\eta}{1-\beta}\right)\left(\frac{c_N^\star}{r}\right)\overline{q}^{\frac{\eta}{1-\beta}} \end{aligned}$$

which is the first half of the desired result. Similarly, if we take the first order condition with respect to q, we obtain

$$0 = -\left(\frac{\eta}{1-\beta}\right) c_N^\star \left(\frac{q}{zk^{\star\beta}}\right)^{\frac{\eta}{1-\beta}} q^{-1} + \alpha\sigma q^{\sigma-1}$$

This implies (multiplying both sides by q) and substituting the definition of $\overline{q}$ that

$$\alpha\sigma q^{\star\sigma} = \left(\frac{\eta}{1-\beta}\right) \times c_N^\star \times \overline{q}^{\frac{\eta}{1-\beta}}$$

Solving for q, we find that

$$q^\star = \left[\left(\frac{\eta/(1-\beta)}{\alpha\sigma}\right) \times c_N^\star \times \overline{q}^{\frac{\eta}{1-\beta}}\right]^{\frac{1}{\sigma}}$$

which completes the proof. □

An immediate result of this proposition is that the identification strategy used in Section 5 is robust to the inclusion of Hicks-neutral capital, up to a change in the interpretation of the returns-to-scale coefficient η, and a change in the interpretation of the span of control z_j.

F Robustness: Narrowly Defined Industries

In this section, I repeat the empirical analysis from Section 3 on a subset of extremely narrowly defined 5-digit industry codes. Following Foster, Haltiwanger, and Syverson (2016), I consider a set of industries that produce a homogenous commodity good, like cement or plywood, and where it is unlikely to be the case that large and small firms have distinct production processes. That is, industries where we can expect that large and small firms are likely completing the same set of tasks.

I choose four industry classifications to examine: sugar cane production and refining, manufacture of plywood, cement manufacturing, and coffee growing/roasting. These industries are chosen to match closely the industries in Foster et al. (2016), for the industries where good analogues exist in the Brazilian CNAE. I define Sugar Production to be any firm with industry codes corresponding to "Cultivation of Sugar Cane" (01139), "Sugar Mills" (15610) and "Sugar Refining and Milling" (15628). Plywood manufacturing consists of a single 5-digit industry code: "Manufacture of Laminated Wood and Plywood, Pressed or Agglomerated Sheets" (20214). Cement manufacturing is a single industry code in the CNAE (26204). Like sugar, I define the coffee industry to consist of several related industry codes: "Coffee Growing" (01325) and "Coffee Roasting and Grinding" (15717), to be consistent with the definition in Foster et al. (2016) (who also aggregate whole beans and roasted/ground coffee mixtures). I do not report results for the remaining industries they consider (Processed and Block Ice, Carbon Black, and Bread) as there do not exist exact analogues for these industries at similar levels of aggregation in the Brazilian CNAE.

Within each of these narrowly defined industries, I repeat the empirical analysis from Section 3: I regress on decile fixed effects the log of the number of occupations within each firm, the average skill level for each dimension of skill, and the within-firm standard deviation of each skill. I report the results for the number of occupations in Table 7. I find in each of the narrowly defined industries a qualitatively similar pattern to what I observe in the aggregate: the number of occupations hired is monotonically increasing in firm size. Even in the comparatively smaller samples, I have extremely tight estimates. For almost all of the deciles, the confidence intervals of the estimates for adjacent deciles are non-overlapping.

I find similar results for the within-firm standard deviation of skills (which I report in Tables 8 to 10). In almost all deciles of all four industries, the dispersion of skills is increasing monotonically, with tight and non-overlapping confidence intervals. This is qualitatively similar to the findings in Section 3, although as with the overall number of occupations, the the size of the increase in dispersion varies across industries. This provides a sharp rejection of the assumption of homotheticity of the production function, even at the 5-digit industry level.

For the average skills (which I report in Tables 11 to 13), I find evidence that for most of the industry-skill pairs, the average skill level is not constant across the firm size distribution. Although we do not see substantial changes in the average cognitive skills in the two agricultural industries (sugar and coffee), this is not entirely surprising. We *do* see substantial variation in these industries for the average level of manual skill across firm sizes, which suggests that to the extent that larger firms hire more specialized workers in these industries, the increase in specialization occurs primarily along the manual dimension of skill, rather than the cognitive. Although the pattern of how average skill varies by firm size is different from industry to industry, it is striking that even within these narrowly defined industries, there is clear evidence that larger firms hire workers with systematically different skills than their smaller counterparts.

	log(Occupations)			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.042*	0.096***	0.059***	0.051***
	(0.024)	(0.030)	(0.009)	(0.019)
deciles: 2	0.169***	0.313***	0.081***	0.102***
	(0.055)	(0.054)	(0.015)	(0.037)
deciles: 3	0.531***	0.484***	0.164***	0.175***
	(0.081)	(0.063)	(0.018)	(0.044)
deciles: 4	0.888***	0.766***	0.277***	0.404***
	(0.083)	(0.064)	(0.020)	(0.046)
deciles: 5	1.433***	0.928***	0.406***	0.566***
	(0.107)	(0.065)	(0.022)	(0.053)
deciles: 6	2.306***	1.184***	0.519***	0.749***
	(0.095)	(0.064)	(0.023)	(0.048)
deciles: 7	2.862***	1.297***	0.665***	1.075***
	(0.099)	(0.068)	(0.023)	(0.049)
deciles: 8	3.023***	1.573***	0.849***	1.283***
	(0.089)	(0.065)	(0.024)	(0.050)
deciles: 9	3.354***	1.990***	1.094***	1.575***
	(0.056)	(0.065)	(0.024)	(0.048)
deciles: 10	3.533***	2.391***	1.819***	2.278***
	(0.064)	(0.088)	(0.032)	(0.063)
N	497	1061	5506	941
R^2	0.857	0.595	0.553	0.732

Table 7: Firm size ranks are calculated using the total quantity of effective labor hired.

	Cognitive w/in Firm Std			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.006*	0.009***	0.007***	0.009**
	(0.004)	(0.003)	(0.001)	(0.004)
deciles: 2	0.026***	0.035***	0.012***	0.009
	(0.009)	(0.007)	(0.003)	(0.007)
deciles: 3	0.059***	0.037***	0.021***	0.027***
	(0.010)	(0.007)	(0.003)	(0.008)
deciles: 4	0.101***	0.052***	0.030***	0.069***
	(0.010)	(0.007)	(0.003)	(0.010)
deciles: 5	0.129***	0.045***	0.042***	0.087***
	(0.010)	(0.006)	(0.003)	(0.011)
deciles: 6	0.149***	0.063***	0.054***	0.119***
	(0.006)	(0.007)	(0.004)	(0.010)
deciles: 7	0.136***	0.056***	0.062***	0.153***
	(0.006)	(0.006)	(0.004)	(0.009)
deciles: 8	0.147***	0.058***	0.071***	0.150***
	(0.006)	(0.005)	(0.004)	(0.009)
deciles: 9	0.149***	0.065***	0.082***	0.163***
	(0.005)	(0.005)	(0.003)	(0.008)
deciles: 10	0.149***	0.074***	0.104***	0.175***
	(0.005)	(0.005)	(0.003)	(0.007)
N	497	1061	5506	941
R^2	0.557	0.132	0.181	0.428

Table 8: Within-firm standard deviation of skills are calculated by as the standard deviation of the skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired.

	Manual w/in Firm Std			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.007	0.010***	0.009***	0.009**
	(0.004)	(0.004)	(0.002)	(0.004)
deciles: 2	0.026**	0.047***	0.014***	0.014**
	(0.010)	(0.009)	(0.003)	(0.007)
deciles: 3	0.093***	0.056***	0.024***	0.026***
	(0.016)	(0.009)	(0.003)	(0.008)
deciles: 4	0.129***	0.073***	0.040***	0.058***
	(0.014)	(0.008)	(0.004)	(0.009)
deciles: 5	0.154***	0.070***	0.049***	0.059***
	(0.011)	(0.007)	(0.004)	(0.008)
deciles: 6	0.193***	0.080***	0.062***	0.078***
	(0.009)	(0.007)	(0.004)	(0.008)
deciles: 7	0.204***	0.073***	0.079***	0.108***
	(0.010)	(0.006)	(0.004)	(0.008)
deciles: 8	0.203***	0.072***	0.089***	0.120***
	(0.008)	(0.006)	(0.003)	(0.008)
deciles: 9	0.211***	0.088***	0.106***	0.137***
	(0.006)	(0.006)	(0.003)	(0.006)
deciles: 10	0.208***	0.096***	0.136***	0.157***
	(0.007)	(0.006)	(0.003)	(0.007)
N	497	1061	5506	941
R^2	0.584	0.148	0.268	0.399

Table 9: Within-firm standard deviation of skills are calculated by as the standard deviation of the skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired.

	Interpersonal w/in Firm Std			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.005	0.011***	0.009***	0.007**
	(0.004)	(0.004)	(0.001)	(0.003)
deciles: 2	0.020**	0.044***	0.012***	0.015**
	(0.008)	(0.009)	(0.003)	(0.006)
deciles: 3	0.053***	0.052***	0.021***	0.023***
	(0.011)	(0.009)	(0.003)	(0.007)
deciles: 4	0.091***	0.071***	0.030***	0.064***
	(0.011)	(0.008)	(0.003)	(0.009)
deciles: 5	0.103***	0.069***	0.043***	0.081***
	(0.009)	(0.008)	(0.003)	(0.009)
deciles: 6	0.141***	0.090***	0.057***	0.115***
	(0.007)	(0.008)	(0.003)	(0.008)
deciles: 7	0.146***	0.082***	0.066***	0.134***
	(0.007)	(0.007)	(0.003)	(0.008)
deciles: 8	0.149***	0.093***	0.077***	0.136***
	(0.007)	(0.006)	(0.003)	(0.007)
deciles: 9	0.145***	0.103***	0.088***	0.141***
	(0.006)	(0.006)	(0.003)	(0.006)
deciles: 10	0.143***	0.116***	0.114***	0.156***
	(0.006)	(0.006)	(0.003)	(0.005)
N	497	1061	5506	941
R^2	0.542	0.201	0.242	0.448

Table 10: Within-firm standard deviation of skills are calculated by as the standard deviation of the skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired.

	Cognitive Skills			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.366***	0.350***	0.340***	0.396***
	(0.027)	(0.012)	(0.008)	(0.022)
deciles: 2	0.045	-0.031**	-0.013	-0.040
	(0.036)	(0.015)	(0.010)	(0.029)
deciles: 3	0.036	-0.050***	-0.039***	0.001
	(0.035)	(0.015)	(0.010)	(0.032)
deciles: 4	0.006	-0.038***	-0.047***	-0.039
	(0.032)	(0.014)	(0.009)	(0.028)
deciles: 5	0.012	-0.064***	-0.045***	-0.028
	(0.031)	(0.014)	(0.009)	(0.029)
deciles: 6	0.004	-0.059***	-0.050***	-0.007
	(0.029)	(0.014)	(0.009)	(0.027)
deciles: 7	-0.002	-0.067***	-0.045***	-0.004
	(0.030)	(0.013)	(0.009)	(0.027)
deciles: 8	-5.496e-04	-0.066***	-0.047***	-0.015
	(0.029)	(0.013)	(0.009)	(0.027)
deciles: 9	-0.007	-0.068***	-0.044***	0.038
	(0.028)	(0.013)	(0.008)	(0.026)
deciles: 10	-0.021	-0.066***	-0.027***	-0.004
	(0.028)	(0.013)	(0.008)	(0.025)
N	497	1061	5506	941
R^2	0.025	0.080	0.016	0.017

Table 11: Average skills are calculated by as the mean of the skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired.

	Manual Skills			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.428***	0.584***	0.524***	0.356***
	(0.028)	(0.018)	(0.008)	(0.020)
deciles: 2	0.083*	-0.003	0.042***	0.053*
	(0.042)	(0.025)	(0.011)	(0.029)
deciles: 3	0.046	0.019	0.045***	0.097***
	(0.039)	(0.022)	(0.011)	(0.027)
deciles: 4	0.045	0.032	0.055***	0.096***
	(0.036)	(0.022)	(0.011)	(0.024)
deciles: 5	0.071**	0.030	0.058***	0.065***
	(0.035)	(0.020)	(0.010)	(0.024)
deciles: 6	0.074**	0.022	0.064***	0.065***
	(0.030)	(0.021)	(0.010)	(0.024)
deciles: 7	0.099***	0.048**	0.069***	0.056**
	(0.029)	(0.020)	(0.010)	(0.025)
deciles: 8	0.121***	0.045**	0.072***	0.066***
	(0.030)	(0.020)	(0.009)	(0.023)
deciles: 9	0.114***	0.043**	0.082***	0.043*
	(0.029)	(0.019)	(0.009)	(0.024)
deciles: 10	0.118***	0.034*	0.082***	0.005
	(0.028)	(0.019)	(0.009)	(0.022)
N	497	1061	5506	941
R^2	0.066	0.021	0.024	0.041

Table 12: Average skills are calculated by as the mean of the skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired.

	Interpersonal Skills			
	Sugar Cane	Plywood	Cement	Coffee
	(1)	(2)	(3)	(4)
(Intercept)	0.396***	0.216***	0.288***	0.391***
	(0.024)	(0.020)	(0.008)	(0.018)
deciles: 2	-0.006	-0.033	-0.026**	-0.041
	(0.032)	(0.025)	(0.010)	(0.025)
deciles: 3	2.131e-04	-0.063***	-0.044***	-0.039
	(0.033)	(0.022)	(0.010)	(0.025)
deciles: 4	-0.030	-0.067***	-0.057***	-0.069***
	(0.029)	(0.022)	(0.009)	(0.025)
deciles: 5	-0.031	-0.078***	-0.060***	-0.040
	(0.028)	(0.021)	(0.009)	(0.025)
deciles: 6	-0.053**	-0.071***	-0.069***	-0.028
	(0.026)	(0.022)	(0.009)	(0.023)
deciles: 7	-0.083***	-0.091***	-0.064***	-0.015
	(0.025)	(0.020)	(0.009)	(0.021)
deciles: 8	-0.094***	-0.095***	-0.070***	-0.032
	(0.026)	(0.021)	(0.009)	(0.022)
deciles: 9	-0.097***	-0.094***	-0.069***	0.021
	(0.025)	(0.021)	(0.008)	(0.020)
deciles: 10	-0.105***	-0.093***	-0.060***	0.001
	(0.025)	(0.021)	(0.008)	(0.020)
N	497	1061	5506	941
R^2	0.118	0.070	0.032	0.031

Table 13: Average skills are calculated by as the mean of the skills of workers hired, weighting by the total quantity of effective labor supplied by each worker. Firm size ranks are calculated using the total quantity of effective labor hired.

G Robustness Check: Leather Working Industry

A key identifying assumption of the model is that all firms within an industry face an identical distribution of tasks. If this assumption fails, then it could be the case that when we observe large firms hiring more occupations, it is because they are expanding the set of tasks that they are doing, rather than hiring more specialized workers to do the same tasks.

To address this potential concern, I re-estimate the model using a more narrowly specified industry industry (Leather working). Because leather goods are relatively homogeneous, we can expect that when we see larger firms hiring more occupations, they really are hiring those additional occupations to do the same set of tasks. I show in this section that the main qualitative results of the paper are robust to this more narrow specification of industry.

I present the estimation results in tables 14 to 16 and fig. 10. In particular, I find a very similar result from the variance decomposition of firm TFP: 31.7% of the variation in firm TFP is due to the fact that more productive firms exogenously choose to hire a more productive mix of workers in production.

In table 17 I present the results from my counterfactual exercise applied to the Leather Industry. My results are qualitatively similar to what I find for the entire manufacturing sector, although I actually estimate larger costs to shutting down the endogenous specialization channel.

	α	β
Cognitive	3.442	1.364
Manual	2.342	1.590
Interpersonal	1.618	2.969

Table 14: Parameter Estimates for Marginal Distributions of $G(\mathbf{x})$

	Absolute Advantage			Comparative Advantage		
	Cognitive	Manual	Interpersonal	Cognitive	Manual	Interpersonal
Cognitive	4.676	-3.460	-1.448	6.776	-0.774	0.030
Manual	-5.175	4.525	0.544	-0.774	0.849	-0.009
Interpersonal	-3.300	-3.108	1.774	0.030	-0.009	8.059e-04

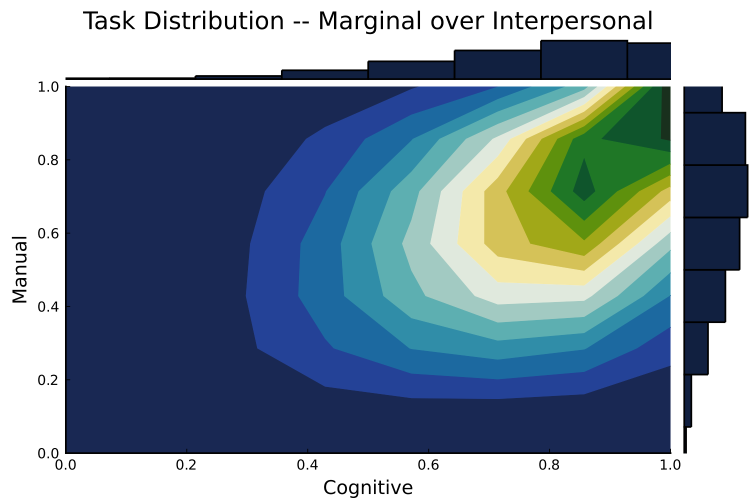
Table 15: Parameter Estimates for the production function parameters A and B, Leather Industry

	Comp	Share
$\text{Var}(\log(z_j \rho_j))$	0.155	—
$\text{Var}(\log(z_j))$	0.019	12.144
$\text{Var}(\log(\rho_j))$	0.087	56.124
$2\ \text{Cov}(\log(z_j), \log(\rho_j))$	0.049	31.732

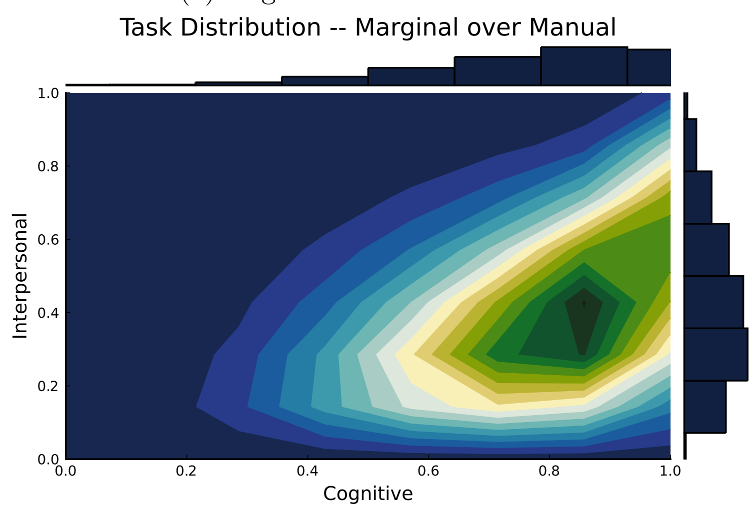
Table 16: Results of the variance decomposition of firm TFP in eq. (21), Leather Industry

	Baseline	$\kappa = 0$	$\kappa = 2 \times \hat{\kappa}$	$\kappa = \text{Large}$
% Δ Consumption		0.928	-0.446	-20.498
% Δ Wage		0.351	-0.200	-7.324
% Δ Output		0.789	-0.379	-17.715
Cognitive	0.367	0.367	0.367	0.362
Manual	0.385	0.386	0.384	0.378
Interpersonal	0.356	0.357	0.356	0.368

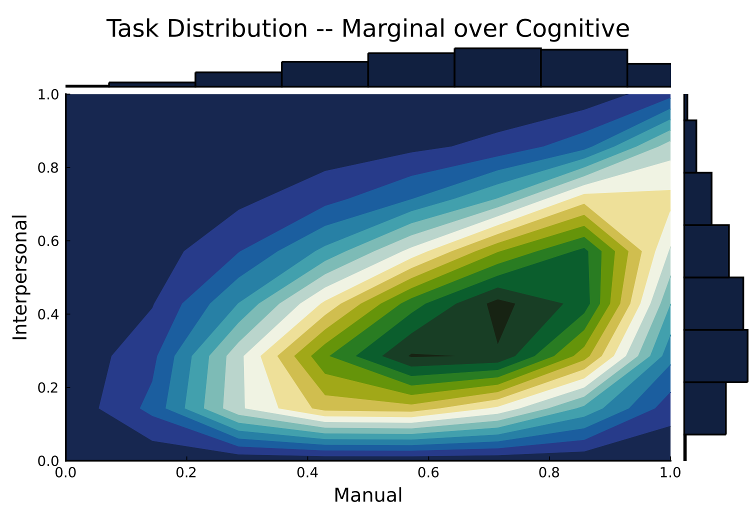
Table 17: Results from Counterfactual Policy Scenarios, Leather Industry



(a) Cognitive vs. Manual Tasks



(b) Cognitive vs. Interpersonal Tasks



(c) Manual vs. Interpersonal Tasks

Figure 10: Estimated Distribution of Tasks, Leather Working, in 2000